

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:25:50 : Search time 6143 Seconds

(without alignments) 5503.423 Million cell updates/sec

Title: US-09-512-581b-2
Perfect score: 7193
Sequence: 1 MAHSTRNDSKITYPPGVK.....OKGRGRPSKTPSPQPKKNV 1391

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 segs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPPO.spool/US09512551/runat_23092003_163547_5486/app_query.fasta_1.1543
-DB-BST -QFMT-fastlap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTWMT-pio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09512551 -ACGN_1_1_4667 -ernat_23092003_163547_5486 -KCPU=3
-NO_MMAPP -LARGESOURCE -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=6 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

ESF:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3787	52.6	3313	11 AK045159	AK045159 Mus muscu
2	3512	48.8	2899	11 AK041682	AK041682 Mus muscu
3	3309.5	46.0	4273	11 AK086753	AK086753 Mus muscu
4	3037	42.2	4427	11 AK032384	AK032384 Mus muscu
5	2170.5	30.2	1359	11 BC032988	BC032988 Mus muscu
6	1395	18.4	926	14 CA980171	CA980171 AGENCOURT
7	1325	18.4	786	14 CA324097	CA324097 UT-M-FY0-
8	1311	18.2	779	14 CA324284	CA324284 UT-M-FY0-
9	1298.5	18.1	786	13 BU703488	BU703488 UT-M-FY0-
10	1295	18.0	782	13 BU363024	BU363024 603790485
11	1285.5	17.9	943	13 BU914579	BU914579 AGENCOURT
12	1269.5	17.6	857	14 CA987660	CA987660 AGENCOURT
13	1262.5	17.6	882	13 BU916668	BU916668 AGENCOURT
14	1239	17.2	931	13 BU914025	BU914025 AGENCOURT
15	1231	17.1	829	13 BX314916	BX314916 BX314916
16	1225	17.0	791	10 BF509252	BF509252 UT-H-B14-
17	1204.5	16.7	771	14 CB312535	CB312535 AGENCOURT
18	1189.5	16.5	748	13 BU957734	BU957734 AGENCOURT
19	1175.5	16.3	905	13 BU940978	BU940978 AGENCOURT
20	1168	16.2	692	12 BM794657	BM794657 K-EST0076
21	1162	16.2	716	14 CB522273	CB522273 UT-M-GH0-
22	1152.5	16.0	876	13 BU912768	BU912768 AGENCOURT
23	1143.5	15.9	734	10 BG256731	BG256731 602370939
24	1142	15.9	679	10 BB094813	BB094813 BB094813
25	1130	15.7	823	13 BU703952	BU703952 UT-M-FY0-
26	1128	15.7	703	14 CB521463	CB521463 UT-M-GH0-
27	1124.5	15.6	868	14 CD301710	CD301710 AGENCOURT
28	1119	15.6	658	14 CD542155	CD542155 B0238804-
29	1108.5	15.4	695	14 CA405800	CA405800 1001949 H
30	1099	15.3	659	12 BU738240	BU738240 603361679
31	1091	15.2	729	10 BG432960	BG432960 602495783
32	1083	15.1	637	14 CB132602	CB132602 K-EST0183
33	1083	15.1	643	13 BU440284	BU440284 604146435
34	1078	15.0	874	13 BU517176	BU517176 AGENCOURT
35	1076	15.0	633	14 CD542034	CD542034 B0237804-
36	1076	15.0	764	13 BU384740	BU384740 603860876
37	1071	14.9	773	9 AL705029	AL705029 DKFP686B
38	1070	14.9	804	13 BU109808	BU109808 603128343
39	1069	14.9	638	14 CD284258	CD284258 G39173.86
40	1067.5	14.8	1142	12 BM466282	BM466282 AGENCOURT
41	1066	14.8	944	10 BF142489	BF142489 601788573
42	1061.5	14.8	848	10 BE873840	BE873840 601483992
43	1061	14.8	657	9 AL846882	AL846882 AL846882
44	1061	14.8	858	13 BU736792	BU736792 AGENCOURT
45	1056.5	14.7	624	14 CB215704	CB215704 NISC_np09

ALIGNMENTS

RESULT 1
LOCUS AK045159 3313 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
enriched library, clone: B130042B12 Product: 49310.1.2
(ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
sequence.
ACCESSION AK045159
VERSION AK045159.1 GI:26337114
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schirml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Wittaker, C., Wilmink, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kotsuk, S., and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 3313)
PUBMED	12108560
REFERENCE	6
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shikata, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Huroko Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/. Location/Qualifiers
FEATURES	1. 3313
SOURCE	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM,DB:BI30042B12"
	/db_xref="taxon:10090"
	/clone="BI30042B12"
	/tissue_type="parthenogenote"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="9.5 days embryo"
CDS	<1..2476
	/note="unnamed protein product; 49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (PRAGMENT) homolog (Homo sapiens) (SPR1096KY4, evidence: FASTV, 92.3%ID, 100%length, match=1430) putative"
	/protein_id="BAC32242.1"
	/db_xref="GI:26337115"
	/translation="VNKSIDGTADDEDEGVPDQAIKAGELIKVLSFTPIFSHAE TPESLACLMDEKVAEALQIFKNTGSEIEDPFRIRALLPVLRHKKKPPROA KVAICHIAIFSSKEVFOAFIEPLHNSLSDPSNDEHITPLVTGHTALAPQFAP KLSLVAVFVIFDLMDNDRLPGKRTYKIMVDEVSPEPTWYQAIKIMVWLLGKNN HSKSGSTLTLTLTIHSDGLDEOGKISRPMSRLTAAGSAIVKLAQDPVCYHEIIT LEQYQCALAINCEQYQVDFVQKLSGRSLRPLEYVIAICALCAKDPKRRRAHA RQCLVKNITVREYIKOHAVASKEKLSLEPEYVPTIHLADPDVYKQDIEQGLD VKCICIMVLEIILAKNNNSHAFIRKVENITKOTKACQGDPTKMKETTYCDVAMN IIMSKSTTYSLESPKDPVLPAREFTOPDNFNTKNTLPPKMSFTTPGPKRTAVLG AVNKPUSASAGKQSTVSSRMETVSNASSSPSPGKIKGLDSSSEHDENDYVMS SPLPGKRSDEKREDPDLVRSLEKPSRKKAPVNDPEKILGMDLTKLVDOKPGKSOR GRRKGRTAQSDSQEQQEERKREKLELLENDQNSPPKGRGRRPPLGCGTSKEP TMTKSSKGNKKLVPPVDDDEEERDIGNTEHRSKSKOHRTSKRAQOQRESSETS VESNOSTPQKGRGRSPKAPSPSPPKTRVGRSKNOVATKENDSSENDVILQASSPVDD TPOBGAEEEDISGVNRRSSKERR"
	polyA_signal
	polyA_site
BASE COUNT	1066 a 709 c 732 g 806 t
ORIGIN	3313
	/note="putative"
	3313
	/note="putative"
Alignment Scores:	7.51e-275
Pred. No.:	3787.00
Score:	97.00%
Percent Similarity:	97.00%
Best Local Similarity:	95.31%
Query Match:	52.6%
DB:	11
	Gaps: 0
US-09-512-581b-2 (1-1391) x AK045159 (1-3313)	
QY	625 ValAsnlySerIleaspGlyThrAlaaspAPspGluaspGluValProthrAspGln 644

|||||
Db 2 GTGAATAATCATGATGGAACACAGATGATGAGAGGTTCACACTATCA 61
Oy 645 AAlaIleargAlaGlyLeuGluLeuLeuLysValLeuSerPheThiHisProIleSerPhe 664
Db 62 GCTATCAGAGGACGGCTTGAGCTGTGAAGGTGTGTCTTTACCCACTTATCTCTTT 121
Oy 665 HisSerIaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysVal 684
Db 122 CACTGTGCTGAGACTTTGAGTCCCTTCTGGCTTGTCTGAAGATGGATGATGAGAAGG 181
Oy 665 AAlaGluAlaAlaLeuGlnIlePheLysasnThrGlySerLysIleGluGluAspPhePro 704
Db 182 GCAGAGAGCTGCACATCAATTTTCAAAAACAGAGAGCAAAATGGAAGAGATTTCCCT 241
Oy 705 HisIleArgSerAlaLeuLeuProValLeuHisSHLysSerLysGlyProProArg 724
Db 242 CACATCAGATCAGCCTTCTCTCTGTACTACATCAAAATCCAAAAAGGACCCTCGG 301
Oy 725 GlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPhe 744
Db 302 CAGGCCAAATACGCATCATTTGTATTCATGCCATATTTTCTAGTAAGAAACCAAGTTT 361
Oy 745 AlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIle 764
Db 362 GCACAGATATTGAGCCCTCGCATAAAGCCTAGATCCAAAGCACCTGGAGCATCTTATA 421
Oy 765 ThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaPro 784
Db 422 ACCCTCTGTGTACTATTGTGCATATTTCTCTTCGACCTGATCAGTTCCTCTCTCT 481
Oy 785 TrpLysSerTrpValAlaThrPheIleValLysAspLeuMetMetAsnAspArgLeuPro 804
Db 482 CTGAAGTCTTTGGTGGCACTTTCTATTGTGAAGGCCCTCTCATATATGACCGCTTCCA 541
Oy 805 GlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetVal 824
Db 542 GGAATAAAGACAACTAAAGCTTTGGSTTCCAGATGAGAGAGTCTCACCCTGAGCAATGGTC 601
Oy 825 LysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSer 844
Db 602 AAAATTCAGGCTATTAAATGATGTGTGATGGCTACTGTGAATGAAAGATATATCAGT 661
Oy 845 LysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGluAspLeu 864
Db 662 AAGTCAGAACTTCACCCCTCAGATTACTAACACGATACCTGCATAGTGAATGGAGTTTG 721
Oy 865 ThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaIaGlySer 884
Db 722 ACAGAACAAAGGAAATTAAGTAACACAGATATGTACAGCCCTGAGACTTGGCTGGAGT 781
Oy 885 AlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlnGluTyr 904
Db 782 GCTATTGTGAAGCTGGCACAGAGGCCCTTACCCAGAGATCATTAACCTGGAGAGATAC 841
Oy 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
Db 842 CAGCTGTGCTATTAAGCATATGATAGTGTATTCAAGTCAGGACAGGTGTTCCCTCAG 901
Oy 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
Db 902 AAACCTTCACAAAGGCCCTTCCGGCTTACGGCTTCCCTTGAATACATGGCCATCTGTCT 961
Oy 945 LeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964
Db 962 CTTTGTGTCCAAAGACCTGTGAAGAGAGGCCACCTGCTAGACACTGTCTGTGTGAAG 1021
Oy 965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
Db 1022 AACATCACTGTGAGAGGAGATACCTGAAGCAGCATGACGCTTTAGTGAATAATTTATG 1081
Oy 985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
|||||

Db 1082 TCCTCTCAGCAGAGTATGTTGTTCCATATACAAATTCACCTTTTGGCACATGACCCAGAT 1141
Oy 1005 TyrValLysValGluAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
Db 1142 TATGTCAAGATACAGATATTGACAACTTAAGAATGTGAAGAATGCCCTTTGGTTGTT 1201
Oy 1025 LeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
Db 1202 CTGGAGATATTATGCTGATAAATGAAAACAAACACCACTGATTTATGAGAAAATGGTA 1261
Oy 1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
Db 1262 GAAATATTAAACACACAAAGATGCTCAGAGCACAGATGATACAAAATGATGAAAAA 1321
Oy 1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
Db 1322 TTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAAGCACCCAGTACAGC 1381
Oy 1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
Db 1382 CTGGAGTCTCCTAAGAGACCCCGTCTGCCAGCTGGTTTTCACCCAGCCTGACAAAGAT 1441
Oy 1105 PheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
Db 1442 TTTAGTAAACACCAAAATTAACCTGCCCTCAGAAATGAATCATTTTTCACCTCTGAAAA 1501
Oy 1125 ProLysThrThrAsnValLeuGluValAlaLysLysProLeuSerSerAlaGlyLysGln 1144
Db 1502 CCTAAACAGCCAAATGTTCTCGACACAGTAATTAAGCCATTTCATCAGCAGCAACAG 1561
Oy 1145 SerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
Db 1562 TCTCAGACCAATATCATCAAGATGGAACCTGAGACACCAAGCAGCAGCTCAACCCA 1621
Oy 1165 SerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsn 1184
Db 1622 AGCTTCTCTGGAAGATCAAGGGAGGCTGTGATCTCTGAAATGATGATCACAGTAAAT 1681
Oy 1185 GluAspTyrThrMetSerSerProLeuProGluLysLysSerAspLysArgAspAspSer 1204
Db 1682 GAAAGATTATACATGTCTTACCTTGGCAGGAAAAAAGACAAAGAGAAAGACCT 1741
Oy 1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgGlyLysThrProValThrGlu 1224
Db 1742 GATCTTGTAGAGTCTGAGTGTGAGAGAACCTGAGAAATGCGGAAAAAACCTCTCAGAGC 1801
Oy 1225 GlnGluGluLysLeuGlyMetAspAspLeuThrThrLysLeuValGlnGluGlnLysProLys 1244
Db 1802 CCTGAAGAAATTAAGTATGATGATGACCTAACTAAGTTGGTACAGAGAACCACTTAAA 1861
Oy 1245 GlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrp 1264
Db 1862 GGCAGTCAAGCGAGCGGAAAAAGAGCCGTAAAGCTTCAAGACTCAGACGAGCAGAGTGG 1921
Oy 1265 ProGluGluLysArgLeuLysGluLysAspIleLeuGlnGluAsnGluAspGluGlnAsnSerPro 1284
Db 1922 CCTGAGAGAAAGAGCACAAGAGAGAGCTCTCGGAAATGAGAGATGAGACAGACGCCCA 1981
Oy 1285 ProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGlu 1304
Db 1982 CCAAAAAAGGCGCAAAAGAGAGCGCCACCAAAACCTTGGTGGGGGAGACATCGAAGGAA 2041
Oy 1305 GluProThrMetLysThrSerLysGlySerLysLysSerGlyProProAlaPro 1324
Db 2042 GAGCCACAAATGAAGAAATCCAAAGAAAGAAACAAAGAAACCTTGTACTCTCTGTAGTA 2101
Oy 1325 GlnGluGlnGluGlnGluGluArgGlnSerGlyAsnThrGlnGlnLysSerLysSerLys 1344
Db 2102 GACGACGATGAAGAAAGAAAGAAACAATTTGGAACACAGAAACATTAAGTCAAAAAAGCAA 2161
Oy 1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIle 1364
Db 2162 CAGCACCCACATCAAAAGAGAGCACACAGAGAGCAATCTCTGAAAAACAAGTGCAGCTT 2221
|||||

```

Qy      1365 GluSerThrGlnSerThrProGlnIlySGlyArgGlyArgProSerIyThrProSerPro 1384
         |||||||
Db      2222 GAATCCACAGACAGTCCACACACAGAAAGACGAGACACATCAAAAGACCATCCCA 2281
Qy      1385 SerGlnProIySylsasnVal 1391
         |||||||
Db      2282 TCACAACCCCCCAAAAAATC 2302

RESULT 2
AK041682
LOCUS   2899 bp      mRNA      linear      HTC 05-DEC-2002
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
            library, clone:A630029M15 product:49J10.1.2 (ANDROGEN-INDUCED
            PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
            (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION AK041682
VERSION   AK041682.1 GI:26334660
KEYWORDS  HTC; CAP trapper.
SOURCE    Mus musculus (house mouse)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS   Carninci, P. and Hayashizaki, Y.
TITLE      High-efficiency full-length cDNA cloning
JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
MEDLINE    99279253
PUBMED     10349636

REFERENCE
AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20499374
PUBMED     11042155

REFERENCE
AUTHORS   Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishitro, T., Harada, A.,
            Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiyagi, K.,
            Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
            Tonedai, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, D.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE      RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
PUBMED     11076861

TITLE
JOURNAL    Nature 409 (6821), 685-690 (2001)
MEDLINE    21085660
PUBMED

REFERENCE
AUTHORS   Arakawa, T., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
            Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
            Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakata, I.,
            Saito, T., Okazaki, Y., Gojoubori, T., Bono, H., Kasukawa, T., Saito, R.,
            Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
            Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
            Knehl, P., Lewis, S., Matsuo, Y., Nakai, I., Pesole, G.,
            Quackenbush, J., Schriml, L. M., Straub, F., Suzuki, R., Tomita, M.,
            Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
            Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, A.,
            Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
            Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
            Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
            Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
            Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
            Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
            Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
            Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
            and Hayashizaki, Y.
TITLE      Functional annotation of a full-length mouse cDNA collection
JOURNAL    Nature 409 (6821), 685-690 (2001)
MEDLINE    21085660
PUBMED

```

```

PUBMED 11217851
REFERENCE
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE      Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
MEDLINE    12053573
PUBMED     12053573

REFERENCE
AUTHORS   Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
            Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
            Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
            Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
            Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
            Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
            Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
            Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
            Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
            Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
            Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
            Muramatsu, M., and Hayashizaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
            URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
            Fax: 81-45-503-9216)
COMMENT    cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of genome exploration research group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL: http://genome.gsc.riken.go.jp/
            URL: http://fantom.gsc.riken.go.jp/
            location/Qualifiers

FEATURES
         source             1..2899
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="FANTOM,DB:A630029M15"
                        /db_xref="taxon:10090"
                        /clone="A630029M15"
                        /tissue_type="thymus"
                        /clone_lib="RIKEN full-length enriched mouse cDNA library"
                        /dev_stage="3 days neonate"
                        117..2222
                        /note="unnamed protein product: 49J10.1.2
                        (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
                        ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo
                        sapiens] (SPR1096KV4, evidence: FASTV, 92.3%ID,
                        100%length, match=1430)
                        putative"
                        /codon_start=1
                        /protein_id="BAC31031.1"
                        /db_xref="GI:26334661"
                        /translation="WVITRNLPDPRKAQDFMKFTQVLEDEKIRKQLEALVSPSC
                        KOAEQVREITKLGNDKOPTNPFLEKIRLIRAPVHIDESTALRKQVKSIC
                        TADDEQGVPTDAIRAGELIVLSFTHSISRSAPETPSILACIMKDEKAYAPAL
                        QIPRTNGSKIEDPPIHRSKALIVLVLRHKSCKGPPROAKTAIFCTIHAIFSKETQPAOI
                        FEPLHKSLDPSNIEHLITPLVTIGIALALAPDPAFAPLKSIVATFVKDILMDRLDGLG
                        KKTITKLWPDDEEVSPEWVAKIQAIKMNRWLKQNNHSGSTSLRLITLILSHSGD
                        LTPQGIKSPDEMRRLRLAAGSAIVKLAQECYIEITLLEQYCALAINDECYQNAV
                        FAQKLRKGSRLRLPLEYVACALCANPKPKERBARAOCVANTIVRREYLRKQHAV
                        SEKULSLPEYVYPTIHLAHDPOYKODIDQLDQVCEGLFVEIILMAKNNESH
                        APTRKAVENIKQTKDAQGPDDTKMNEKLTVCYVAMNITSKSTTSLSPKRPVLP
                        RFTTQDPKNSNTKNYLPPEMKSFPTPGKFNATNVGAVKPLSSAGKQSQTSRME
                        TVSNASSNPSSNGRIKGRLDVSEMDHSENEIDYTWSSPLPGKRSKDRPDLVLRVC

```


BASE COUNT 868 a 582 c 617 g 832 t
ORIGIN LVGRVRLIVLVIPAL"
Alignment Scores:
pred. No.: 3,51e-254 Length: 2899
Score: 3512.00 Matches: 683
Percent Similarity: 98.99% Conservative: 2
Best Local Similarity: 98.70% Mismatches: 7
Query Match: 48.83% Indels: 0
DB: 11 Gaps: 0
US-09-512-581b-2 (1-1391) x AK041682 (1-2899)
QY 517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
DB 78 CAGACAGATGCCAGTGTCAAGGCCATATTTCTAAAGTATGTTATCCACAAATTTTG 137
QY 537 ProAspProGlyLysAlaIleAspPheMetLysLysPheThrGlnValLeuGluAspAsp 556
DB 138 CCAGATCCCTGCTAAGGCCCAAGACCTTCATGAAGAAGTTCCACAAAGTTGGAAAGATGAT 197
QY 557 GluLysIleArgLysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAla 576
DB 198 GATGAATAATGGAACAGCTGGAAGCCCTTGTTAGTCCACATGTTCCGCAAGCGGCT 257
QY 577 GluGlyCysValArgGluIleThrLysLysLeuGlnAspProLysGlnProThrAspPro 596
DB 258 GAAGGCTGTGCTGTAATTAAGAGTTGGGCAACCTTAAGCAGGCTTCAAAACCT 317
QY 597 PheLeuGluMetCileLysPheLeuLeuGlnArgIleAlaProValHisIleAspThrGlu 616
DB 318 TTTCTGGAATATCAAGTTCTCTGAGAGATGCTGCTGACATAGACACTGAA 377
QY 617 SerIleSerAlaIleuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGlu 636
DB 378 TCCATAGTGTCTTATCAAGCAAGTGAATTAATCAATGATGGAACACAGATGATGAG 437
QY 637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGlnLeuLeuLysValLeu 656
DB 438 GATGAGGGTGTCCAACTGATCAGGTATCAGGGCAGGCTTGAGTGTGTAAGGCTG 497
QY 657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
DB 498 TCTTTTACGACCCATCTCTCTTCACTCTGAGACGTTGAGCTCCCTGCGCTGT 557
QY 677 LeuLysMetAspAspGluLysValAlaGlnAlaIleLeuGlnIlePheLysAsnThrGly 696
DB 558 CTGAAGATGATGATGAGAAGGTGGCAGAAGCTGCATCAAAATTTTCAAAAACACAGGA 617
QY 697 SerLysIleGluGlnAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
DB 618 ACCAAATTTGAAGAGATTTCCCTCAGATCAGATCAGCTTGCTTCTCTACTACTAC 677
QY 717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
DB 678 AATTCAAAAAGGACCACTCGCAGGCGCAATACGCCATCATGTGATTCATGCGCATA 737
QY 737 PheSerSerLysGlnThrGlnPheAlaGlnIlePheGlnProLeuHisLysSerLeuAsp 756
DB 738 TTTTCTAGTAAGAAACCCAGTTTGACAGATATTGAGCCCTCTGATAAAGCCCTAGAT 797
QY 757 ProSerAsnLeuGlnHisIleLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeu 776
DB 798 CCAACCAACCTGGAGCATCTTATTAACCCCTGCGACATATGGTCATATTCCTCTTC 857
QY 777 AlaProAspGlnPheAlaAlaProTrrPlySerTrrPvalAlaThrPheIleValLysAsp 796
DB 858 GCACCTGATCAAGTGTGCTGCTCTGGAAGCTTTGGTGCAAACTTTCATGTGGAAGGAC 917
QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrPalaProAspGlu 816
DB 918 CTCTCTATGAAATGACCGGCTTCCAGGAATAAAGACAACTAAAGCTTTGGGTTCCAGATGAG 977

QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
DB 978 GAAGTCTCACCTGAGACATGCTCAAAATTCAGGCTATTAATATGATGTTGCGATGGCTA 1037
QY 837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThr 856
DB 1038 CTTGGAAATGAAATATATCAGATGAAGTCCAGAACTTCCACCCCTCAGATTAATCAACAG 1097
QY 857 IleLeuHisSerAspGlyAspLeuThrProGluGlnGlyLysIleSerLysProAspMetSer 876
DB 1098 ATACTGCATAGATGAGGATTTTGACAGAACAGGAAATTAAGTAACCAATATATGCTCA 1157
QY 877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTrrHis 896
DB 1158 CGCCTGAGACCTTGCTGCTGGAGTGTATTTGGAAGCTGCGACAGAGGCCCTTATCCAC 1217
QY 897 GluIleIleThrLeuGluGlnTrrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
DB 1218 GAGATCATTAACCTGGAGCAGTACCAAGCTGTGTCATTAAGCATCATGATGATGATGAT 1277
QY 917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgPro 936
DB 1278 CAAGTTCAGGAGAGTGTGCTCAGAACTTCAAAAGGCTTCCCGCTTACGGCTTCCC 1337
QY 937 LeuGluTrrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
DB 1338 CTTGAGTACATGGCCATCTGTGCTCTTGTGCAAAAGACCTTGAAAGAGAGGAGGCC 1397
QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTrrLeuLysGlnHis 976
DB 1398 CATGCTAGACAGTGTGGTGAAGAACATCACTGTGAGGAGGAGTACTCTGAAGCAGCAT 1457
QY 977 AlaAlaValSerGluLysLeuSerLeuLeuProGluTrrValValProTrrThrIle 996
DB 1458 GCAGCTGTAGGAAATATATGTCTCTTCAACAGATATGTTGTTCCATATCAATT 1517
QY 997 HisLeuLeuAlaHisAspProAspTrrValLysValGlnAspIleGluGlnLeuLysAsp 1016
DB 1518 CACCTTTTGGCAGATGACCAGATATATGCAAAAGTACAGAGATATTGAACCACTTAAGAT 1577
QY 1017 ValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
DB 1578 GTGAAGAATGCTTGTGTTGTTCTTGCGAATATGATGCGCTTAAATGAAACAAACAGC 1637
QY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
DB 1638 CATGCATTTATCAGAAATAATGATGAATAATTAACACAGCAAAAGATGCTCAAGGACCA 1697
QY 1057 AspAspAlaLysMetAsnGluLysLeuTrrThrValCysAspValAlaMetAsnIleIle 1076
DB 1698 GATGATACAAAATATGATTAATAATTTGTACACGATGATGATGATGATGATGATGATGAT 1757
QY 1077 MetSerLysSerThrThrTrrSerLeuGlnSerProLysAspProValLeuProAlaArg 1096
DB 1758 ATGTCCAAAGACGACACGATGACGCTTGAGTGTCTTAAAGGACCCGCTGTGCGACCTGG 1817
QY 1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrrLeuProProGluMet 1116
DB 1818 TTTTTCACCCAGCTGACAGAAATTTTATGATACACCAAAATTAATTAATTAATTAATTAAT 1877
QY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlnAlaValAsnLys 1136
DB 1878 AATATCATTTTTCACCTCGGAAACCTTAACACGCAATGTTCTCGGAGACGATTAATAG 1937
QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
DB 1938 CCACCTTATCATCGCAGGAGAAACAGTCTCAGACCAATATCATCAAGATGAGAACTGTGAGC 1997
QY 1157 AsnAlaSerSerSerSerAsnProSerSerProGlyArgGlyIleLysGlyArgLeuAspSer 1176
DB 1998 AACGCAAGCAGAGCTCCACCAACCAAGCTCTCTGGAAGATTCAAAGGAGGCTTGTATAGC 2057

Oy 1177 SerGluMetAspHisSerGluAsnGluAspTyrThrMetSerProLeuProGlyLys 1196
 |||||||
 Db 2058 TCTGAATGATGCACAGTGAATAATGATATACATGCTTCACCTTCACAGAA 2117
 |||||||
 Oy 1197 LysSerAspLysArgAspAspSerAspLeuValarg 1208
 |||||||
 Db 2118 AAAAGTGCAAGAGAGAACCCCTGATCTTGTAAAG 2153
 |||||||
 RESULT 3
 AK086753
 LOCUS
 DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
 library, clone:0930049H23 Product:49310.1.2 (ANDROGEN-INDUCED
 PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
 (FRAGMENT) homolog [Homo sapiens], full insert sequence.
 ACCESSION AK086753
 VERSION AK086753.1 GI:26103739
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL MEDLINE 20499374
 PUBMED 11042159
 REFERENCE
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, T., Itoh, M.,
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, S.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Okada, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed capillary sequencer
 JOURNAL MEDLINE 20530913
 PUBMED 11076861
 REFERENCE
 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L.M., Stabill, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Kono, H.,
 Baldarelli, R., Barish, G., Blake, J., Boffelli, D., Bujunga, N.,
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Butt, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D.A., Kamliya, M., Lee, N.H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
 Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
 Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 JOURNAL MEDLINE 21085660
 PUBMED 11217851

REFERENCE
 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 4273)
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numaizaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshitake Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/.
 FEATURES
 source
 1. 4273
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM,DB:D930049H23"
 /db_xref="taxon:10090"
 /clone="D930049H23"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="15 days embryo"
 <1. 2245
 /note="49310.1.2 (ANDROGEN-INDUCED PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog
 [Homo sapiens] (SPTR1096KV4, evidence: FASTV, 92.3%ID,
 100%length, match=1430)
 putative"
 misc_feature
 BASE COUNT 1363 a 840 c 899 g 1171 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-238 Length: 4273
 Score: 3309.50 Matches: 652
 Percent Similarity: 96.24% Conservative: 14
 Best Local Similarity: 94.22% Mismatches: 23
 Query Match: 46.01% Indels: 5
 Gaps: 1
 US-09-512-581b-2 (1-1391) x AK086753 (1-4273)
 Oy 702 AspheProHisIleArgSerAlaLeuLeuProValLeuHisIstLysSer-LysLysG 721
 |||||||
 Db 1 GATTTCCCTGCACATGCATGACGCTTCCTCTACTACATGCACAAATCCAAAGAG 60
 |||||||
 Oy 721 yProPArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLysG 741
 |||||||
 Db 61 ACCACCTGGGCGGCAATGACCAATGCATGCTGATGATGCTATTTCTAGTAAGA 120
 |||||||

QY 741 uThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuG 761
 |||||
 Db 121 AACCCAGTTGGACAGATATTGGAGCTCTGCATTAAGCCTTAGATCCAGCAAGCTGGA 180
 QY 761 uHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPh 781
 |||||
 Db 181 GCATCTTAATACCCCTCTGTCACATATGTCATATGCTCTTCGCAAGCTGATGAGTT 240
 QY 781 eAlaAlaProTrrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 801
 |||||
 Db 241 CCGTCTCCCTGCAAGTCTTGGTGGCACTTTCATGTCGAAGGAGCCCTCTCAAGATGA 300
 QY 801 pArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlnGluValSerProG 821
 |||||
 Db 301 CCGGCTCCAGGAAAAAACAACAACCTTAGCTTGGGTTCCAGATGAGGAAGCTCTCACCTTA 360
 QY 821 uThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAs 841
 |||||
 Db 361 GACAATGGTCAAAATTCAGGCTAATTAAATGATGGTTCGATGGCTACTGGAAATGAAGA 420
 QY 841 nAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAs 861
 |||||
 Db 421 TTAATCAAGTAAAGTCAGAACTCCAGCTCAGATTACTAACAACGATACCTGCATAGTGA 480
 QY 861 pGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAl 881
 |||||
 Db 481 TGGGATTTGACAGAAACAAGGAAATATGTAACACAGATATGTCCAGGCTGAGACTTGC 540
 QY 881 eAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLe 901
 |||||
 Db 541 TCTCTGGAGTGTATTGTGGAACCTGCGACAGAGCCCTGTACACAGATCATATTCACACT 600
 QY 901 uGlnGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnVa 921
 |||||
 Db 601 GGAGCAGTACCAAGCTGTGTCATTCACCTCAATGATGATGATTAATCAAGTCAAGGCAAG 660
 QY 921 lPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu - ProLeuGluTyrMetA 941
 |||||
 Db 661 GTTCGCTCAGAAACTTCACAAAGGCTTCCCGCTTACGCGTTCCTCCCTGGAGTACATGG 720
 QY 941 lAlleCysAlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaArgGlnC 961
 |||||
 Db 721 CCATCTGTGCTCTTGTGTCAAAGACCCTGTGAAGAGAGGAGGAGCCATGCTAGACACT 780
 QY 961 yLysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerG 981
 |||||
 Db 781 GCTTGTGTAAGAACAATCCC -GTGAGGAGGAGGAGTACCTGAAGAGCTGAGCTGTATAGT 839
 QY 981 lLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
 |||||
 Db 840 AAAAATTAATGTCTCTTACCAAGATATGTTCCATATACATTCACCTTTGGGCAC 899
 QY 1001 iAspProAspTrrpValLysValGlnAspIleGlnLeuLysAspValLysGluCysL 1021
 |||||
 Db 900 ATGACCACGATTAATGTCAAAAGTACAGGATTAATGAACAACCTTAAGATGGAAGAATGCC 959
 QY 1021 eutTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleA 1041
 |||||
 Db 960 TTTGGTTTGTCTGAGATATTGATGGCTAAAAAATGAACAAACAGCCATGATTTATGA 1019
 QY 1041 rGlyMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysM 1061
 |||||
 Db 1020 GAAAAATGCTGAAAAATATTAAACAGACAAAGATGCTCAAGAGACAGATGATACAAAAA 1079
 QY 1061 eAsnGlnLysLeuTrrpValLysAspValAlaMetAsnIleIleMetSerLysSerT 1081
 |||||
 Db 1080 TGAATGAATAATGTACACCGTGTGATGTTGCCATGAACATCAATCAATGATCCAAAGACA 1139
 QY 1081 hTrpThrTyrSerLeuGlnSerProLysAspProValLeuProAlaArgPhePheThrGlnP 1101
 |||||
 Db 1140 CCACGTACAGCTGTGAGTCTCTTAAGAGACCCCGTGTGCTGCAGCTGGTTTTCACCCACAC 1199
 QY 1101 roAspLysAsnPheserAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheT 1121

Db 1200 CTGACAGAAATTTTAACTAACACCAAAAATTTACTGCTGCACAAATGAATATATTTC 1259
 QY 1121 hTrpGlyLysProLysThrThrAsnValLeuGlyValAlaLysLysProLeuSerSera 1141
 |||||
 Db 1260 CTCCTGGAAGAACTTAAACAGCCAAATGTTCTCGAGCAGTATTATAAGCCACTTTCAG 1319
 QY 1141 lAGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAlaLaserSers 1161
 |||||
 Db 1320 CAGGCAAAAGTCTCAGACCAATATCAAGAATGGAAACTGTGACCAACGCAAGCAGCA 1379
 QY 1161 eSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
 |||||
 Db 1380 GCTCCACCCCAAGCTCTCTCGAAGAGATAG -GGAGGCTTGATACCTTGAAATGAGATC 1438
 QY 1181 iSerGluAsnGluAspTrrpThrMetSerSerProLeuProGlyLysLysSerAspLysA 1201
 |||||
 Db 1439 ACAGTGAATAATGATATATACAAATGCTTTCACCTTTCACAGAAAAAAGTACAGGA 1498
 QY 1201 rGAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrP 1221
 |||||
 Db 1499 GAAAGAGCCCTATCTTGTAAAGTCTGAGTTGGAAGAGCTTGAAGTCGGAAGAAAAAGCA 1558
 QY 1221 roValThrGlnGlnGlnGlnLysLeuGlyMetAspAspLeuThrLysLeuValGlnGln 1241
 |||||
 Db 1559 CTGTCAAGACCCCTGAAGAGAAATTAAGTATGATGAGACTTAAGTTGTTGATACAGAAC 1618
 QY 1241 lLysProLysGlySerGlnArgSerArgLysArgLysIleThrAlaSerGluSerAspG 1261
 |||||
 Db 1619 AGAAACCTTAAGAGGAGTGCAGGAGCCGGAAGAGGCGCTTCACCTTCAGCTCAGACG 1678
 QY 1261 lGlnGlnTrrpProGlnGlnLysArgLeuLysGluAspIleLeuGluAsnGluAspGln 1281
 |||||
 Db 1679 AGCAGCAGTGGCTGTAGAGGAAAGAGGACCAAAAGAGAGCTCTTGAAAAATGAGATGAC 1738
 QY 1281 lAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyT 1301
 |||||
 Db 1739 AGAAGACCCCAACAAAAAGGCAAAAGAGGAGGAGCCCAACCTTGTGTGGGGGA 1798
 QY 1301 hTrpLysGlnGluProThrMetLysThrSerLysLysGlySerLysSerGlyP 1321
 |||||
 Db 1799 CATCGAAGGAAAGAGCCCAAAATGAAGAAACATCCAAAGAAAGAAACAAAGAAACCTTGTGAC 1858
 QY 1321 roProAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1341
 |||||
 Db 1859 CTCCTGTACTACACGACGATGAGAACAAAGAAAGCAAAATTTGAAACACAGAACATACT 1918
 QY 1341 eLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGln 1361
 |||||
 Db 1919 CAAAAAGCAAAACAGACACGACATCAAAAGAGAGCACACAGAGACAGAAATCTCTGAAA 1978
 QY 1361 eSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyLysArgLysArgProSerLysT 1381
 |||||
 Db 1979 CAAGTCAAGTTAAATCCACACAGTCCACACCAACAAAGAGAGAGGAAAGACATCAAAAG 2038
 QY 1381 hTrpSerProSerGln---ProLysLysAsn 1390
 |||||
 Db 2039 CACCATCACCATCACACACCCCAAAAAAAT 2070
 RESULT 4
 AK032384 4427 bp mRNA linear HTC 05-DEC-2002
 LOCUS AK032384
 DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
 enriched library, clone:6430510p15 product:49310.1.2
 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
 PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
 sequence.
 AK032384
 ACCESSION AK032384.1 GI:26082814
 VERSION
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi. Muridae; Murinae; Mus.
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to
MEDLINE	prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
TITLE	Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
JOURNAL	Sunil, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
MEDLINE	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
PUBMED	Fujiwaka, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matshiki, M.,
REFERENCE	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
AUTHORS	Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format
JOURNAL	sequencing pipeline with 384 multicapillary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PUBMED	11076661
REFERENCE	4
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
TITLE	Arkawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
JOURNAL	Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahata, I.,
MEDLINE	Salto, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Salto, R.,
PUBMED	Kadota, K., Matsuda, H., Asburner, M., Batalov, S., Casavant, T.,
REFERENCE	Flisbacher, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
AUTHORS	Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, T., Peele, G.,
TITLE	Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
JOURNAL	Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
MEDLINE	Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
PUBMED	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
REFERENCE	Flatcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
AUTHORS	Hofmann, M., Hume, D.A., Kamliya, M., Lee, N.H., Lyons, P.,
TITLE	Montchioni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
JOURNAL	Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
MEDLINE	Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
PUBMED	Toyo-oka, K., Wang, K.H., Weitz, C., Wiltaker, C., Wilmberg, L.,
REFERENCE	Wyszaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
AUTHORS	and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research
TITLE	Group Phase I & II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation
MEDLINE	of 60,770 full-length cDNAs
PUBMED	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 4427)
AUTHORS	Adachi, J., Aizawa, K., Akiyama, T., Arkawa, T., Bono, H., Carninci, P.,
TITLE	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
JOURNAL	Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirose, T.,
MEDLINE	Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
PUBMED	Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
REFERENCE	Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
AUTHORS	Nakamura, M., Nishi, K., Nomura, K., Nakazaki, R., Ohno, M., Ohnato, N.,
TITLE	Okazaki, Y., Salto, R., Salto, H., Sakai, C., Sakai, K., Sakazume, N.,
JOURNAL	Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
MEDLINE	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
PUBMED	Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
REFERENCE	Muramatsu, M. and Hayashizaki, Y.
AUTHORS	Direct Submission
TITLE	Submitted (15-JUL-2001) Yoshihide Hayashizaki, The Institute of

COMMENT	FEATURES	SOURCE
Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	cdna library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://phantom.gsc.riken.go.jp/ Location/Qualifiers	1. 4427 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6j" /db_xref="FANTOM.DB:6430530P15" /db_xref="taxon:10090" /clone="6430530P15" /sex="male" /tissue_type="Olfactory brain" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" <1. 2062 /note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHOTOF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens) (SPR1Q96KV4, evidence: FASTV, 92.3%ID, 100% length, match=1430) 4406. 4411 putative" polyA_signal 4406. 4411 /note="putative" polyA_site 4427 /note="putative" BASE COUNT 1423 a 832 c 919 g 1253 t ORIGIN
Alignment Scores:		
Score: No.:	3.96e-218	Length: 4427
Percent Similarity:	3037.00	Matches: 591
Best Local Similarity:	95.888	Conservative: 14
Query Match:	93.668	Mismatches: 24
DB:	42.228	Indels: 3
	11	Gaps: 1
US-09-512-581B-2 (1-1391) x AK032384 (1-4427)		
OY 761 G1UH1S1eU1eThrProLeuVal1hrl1eGly1S1leAla1leU1eAla1ProAsp1n 780		
Db 3 GACCACTCTATAAACCCCTGCTGCACATATGGTCTATATGGCTCTTCGACCTGATCAG 62		
OY 781 PheAlAlaProTrpPrySerTrpValAlaThrPheIleVal1LysAsp1leU1eMetAsn 800		
Db 63 TTGGCTGCTCTCTGGAAGCTTTGGTGGCACTTTCATTGTAAGAGCACTCTCATGAAAT 122		
OY 801 AspArg1eUProG1Lys1LysThrThr1Lys1eU1rPa1ProAsp1LuG1Val1SerPro 820		
Db 123 GACCGGCTTCACAGAAAGAAAGACACATGAGCTTGGTTCAGATGAGGAAGTCTCACCT 182		
OY 821 G1UThMeVal1Lys1LeG1nAlAl1eLysMe1eVal1ArgTrp1eU1eGlyMe1Lys 840		
Db 183 GAACACATGGTCAAAATTTAGGCTATTAAATGATGAGTTCGATGGCTACTTGAATGAAG 242		
OY 841 AspAsnHisSer1LysSerG1Yh1SerThr1eU1rPa1Arg1eU1eU1rThr1IleU1HisSer 860		
Db 243 AATATACACAGTAAAGTCACAGACTTCACACCTTCAGATTAACAAAGATATCTGATAGT 302		
OY 861 AspG1YAsp1eU1hTrG1LuG1nG1Lys1Lys1eSer1YProAspMetSerArg1eU1rG1eU 880		
Db 303 GAAGGGAGATTTCACAGAACAGGAAATTTGTAATCAACAGATATGTGCAGCCTGAGACATT 362		

QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlnIleIleThr 900
 |||||
 Db 363 GCTGCTGGAGAGTGCATTGTTGAAGCTGGACAGAGACCCTGTTACACAGATCATTACA 422
 QY 901 LeuGluGlnIleTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
 |||||
 Db 423 CTGGAGCAGTACACCTGTGTGCATTAGCCATCAATGATGAGTATCATCAAGTCAGAGCAG 482
 QY 921 ValPheAlaGlnIleLysLeuHisIleLysGlnLeuSerArgLeuArgLeuProLeuGluTyrMet 940
 |||||
 Db 483 GGTTCGTGCACAAACTTCACAAAGGCCCTTCCCGCTTACGGCTTCCTCCCTGAGTACATG 542
 QY 941 AlaIleCysAlaLeuCysAlaIleLysAspProValLysGlnArgArgAlaHisAlaArgGln 960
 |||||
 Db 543 GCCATCTGCTGCTTGTGTCACAAAGCCCTGTGAAGAAGAGGAGGCCATCTGTACAGC 602
 QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
 |||||
 Db 603 TCCCTGGTGAAACATCACTGTGAGGAGGAGTACCTGAGCAGCATCCTGCTTACT 662
 QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuAla 1000
 |||||
 Db 663 GAAATAATTAATCTCTCTTACCAGAGTATGTTGTCATATACAAATTCACCTTTGGCA 722
 QY 1001 HisAspProAspTyrValLysValGlnAspIleGlnGlnLeuLysAspValLysGluCys 1020
 |||||
 Db 723 CATGACCCAGATTATGTCAAGTACAGATATTGAACAACCTTAAGATGTGAAGAAGATGC 782
 QY 1021 LeuTyrPheValLeuGluIleLeuMetAlaLysAsnGlnAsnSerHisAlaPheIle 1040
 |||||
 Db 783 CTTTGGTTGTTCTGTGAGATATTGATGGCTAAATAAGAAACACGCCATCATATTATC 842
 QY 1041 ArgLysMetValGlnAsnIleLysGlnIleThrLysAspAlaGlnGluProAspAlaLys 1060
 |||||
 Db 843 ACAAAAGGTGTAAGAAATTAATAACAGACAAAGATGCTCAAGAGCACATATATACAA 902
 QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
 |||||
 Db 903 ATGAATGAAATAATTGTACACCGCTGTGTGATGTTGCCATGACATCATCTGTCCAAAGAC 962
 QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGln 1100
 |||||
 Db 963 ACCAGTACAGCCTGTGAGTCTCTCAAGACCCCGTGTCTCCAGCTCGGTTTTCACCCAG 1022
 QY 1101 ProAspLysAsnPheSerAspThrLysAsnTyrLeuProGluMetLysSerPhePhe 1120
 |||||
 Db 1023 CCTGACAGAAATTTAGTACACCAAAATTAACCTGCTCCAGAAATGAATCATTTTTC 1082
 QY 1121 ThrProGluLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
 |||||
 Db 1083 ACTCCTGGAAACCTTAACACCCCAATGTTCTCGAGCAGCT- AATTAAGCCACTTTCATCA 1141
 QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
 |||||
 Db 1142 GCAGCAAAACACTCAGACCAATCATCAAGAAATGAGAAATGTGAGCAACCAAGCAGC 1201
 QY 1161 SerSerAsnProSerSerProGluArgIleLysGlnArgLeuAspSerSerGluMetAsp 1180
 |||||
 Db 1202 AGCTCCAAACCCAGAGCTCTCTGGAAGAGTCAAGGGGAGGCTTGATAGCTCTAATAATGAT 1261
 QY 1181 HisSerGluAsnGlnAspTyrThrMetSerSerProLeuProGluLysLysSerAspLys 1200
 |||||
 Db 1263 CACCATGAAATAAGAAATTAATACATGCTTCACTTCCAGGAAAAAGTGCACAG 1321
 QY 1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGluArgLysLysThr 1220
 |||||
 Db 1322 AGAGAGAGACCCGATCTT-----TCTGAGTTGAGAGAACCTTAGAAGTCGGAAGAAAGCA 1375
 QY 1221 ProValThrGlnGlnGlnLysLeuGlnGlnMetAspAspLeuThrLysLeuValGlnGlu 1240
 |||||
 Db 1376 CCTGTGCACAGCCCTGAAGAGAAATTAAGTATGATGACCTAAGTAAAGTTGGTACAGGAA 1435
 QY 1241 GlnLysProLysGlnSerGlnArgSerArgLysArgGlnHisThrAlaSerGluSerAsp 1260

Db 1436 CAGAAACCTTAAGCAGTACGAGGCCGGAAGAGGCCCTACAGCTTACAGCTCAGAC 1495
 |||||
 QY 1261 GlnGlnGlnIleTyrProGluGlnLysArgLeuLysGlnAspIleLeuGlnLysAspGlu 1280
 |||||
 Db 1496 GAGCAGAGTGGCTGTGAGAGAGAGAGCACAAGAGAGAGACTCTCTGAAATGAGATGAG 1555
 QY 1281 GlnAsnSerProProLysGlnLysArgGlyArgProProLysProLeuGlnLysGly 1300
 |||||
 Db 1556 CAGAACAGCCACCAAAAAAGGCAAAAGAGCAGGCCACCAAAACCTTGTGTGGGGG 1615
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlnSerLysLysSerGly 1320
 |||||
 Db 1616 ACATGGAAGGAAGAGCCCAATGAAACATCATCAAGAAAGAAACAGAAAAACTTGTGA 1675
 QY 1321 ProProAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1340
 |||||
 Db 1676 CCTCCTGTAGTACAGCAGCATGAAAGAAAGAAACAAATTTGAAACACAGACATTAAG 1735
 QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
 |||||
 Db 1736 TCAAAAGCAACAGCAGCCGACATCAAGAGACACACAGACAGCAATCTCTGAA 1795
 QY 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
 |||||
 Db 1796 ACAAGTGCAGTTGAATCCACACAGTCCACACAGAAAGAGAGAGCAATCATCAAA 1855
 QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
 |||||
 Db 1856 GCACCATCATCATCAACACCCCAAAAAAATC 1888
 |||||
 RESULT 5
 BC032988 1359 bp mRNA linear HTC 20-SEP-2002
 LOCUS BC032988 Mus musculus, clone IMAGE:1179696, mRNA.
 DEFINITION BC032988
 ACCESSION BC032988.1 GI:21426933
 VERSION HTG.
 KEYWORDS Mus musculus (house mouse)
 SOURCE
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 (bases 1 to 1359)
 REFERENCE
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Marcello Bento Soares, Ph.D.
 CDNA Library Preparation: Soares Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 Contact: amadanesystemsbiology.org
 Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clon distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRK Plate: 66 Row: C Column: 13
 This clone has the following problem: no polyA-tail.
 Location/Qualifiers
 1..1359
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:1179696"
 /tissue_type="Mammary gland"

FEATURES

source

10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
 Cloned directionally, priming method: oligo-dt. cDNA
 enrichment: >1k bp. Average insert size 1.8k bp. Priming
 sequence: 5'GACATGCTTTCATGATCGGAGCGGCCGCC(7) 3'. Tissue
 contributed by, David Rowe. Library constructed by Resgen,
 Invitrogen Corp."

BASE COUNT 332 a 199 c 222 g 172 t 1 others

Alignment Scores:

Pred. No.: 3.98e-95 Length: 926
 Score: 1395.00 Matches: 277
 Percent Similarity: 94.02% Conservative: 6
 Best Local Similarity: 92.03% Mismatches: 16
 Query Match: 19.39% Indels: 3
 DB: 14 Gaps: 0

US-09-512-581b-2 (1-1391) x CA980171 (1-926)

QY 1015 LysAspValLysGluCysLeuTrpPheValLeuGluLeuMetAlaLysAsnGluAsn 1034
 |||||||
 Db 26 AAAGATGTGAAGAAATGCTTGGTTGTTCTGAGATATTGATGGCTAAATGAAGAAC 85
 |||||||
 QY 1035 AsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGln 1054
 |||||||
 Db 86 AACAGCATGATTTATCAGAAAAATGTAAGAAATATTAAACAGACAAAGATGCTCAA 145
 |||||||
 QY 1055 GlyProAspAspAlaLysMetAsnGluLysLeuThrValCysAspValAlaMetAsn 1074
 |||||||
 Db 146 GGACCGATGATGACAAATGAATGAAATGTAATGTAACCGTGATGATGTCGATGAC 205
 |||||||
 QY 1075 IleIleMetSerLysSerThrThrTrpSerLeuGluSerProLysAspProValLeuPro 1094
 |||||||
 Db 206 ATCATCATGTCCAAAGACACACGATACAGCTGAGATCTCTCTAGAGACCCCGTGTGCCA 265
 |||||||
 QY 1095 AlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrpLeuProPro 1114
 |||||||
 Db 266 GCTGGGTTTTCACCCAGCTGACAGAAATTTTAAACACCAAAATTAACGCTGCCA 325
 |||||||
 QY 1115 GluMetLysSerPhePheThrProGlyLysProLysThrAsnValLeuGlyAlaVal 1134
 |||||||
 Db 326 GAATGCAATCATTTTTCACCTCTCGTGAACCACTAAACGCCATGTTCTCGAGCAGTT 385
 |||||||
 QY 1135 AsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
 |||||||
 Db 386 AATTAACCCACTTTCATCAGCAGCAACAGCTCTCAACCAATCATCAAGATGGAAC 445
 |||||||
 QY 1155 ValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeu 1174
 |||||||
 Db 446 GTGAGCAACGCAAGCAGCTCCACCAAGCTCTCGAAGATCAAGGGGAGGCTT 505
 |||||||
 QY 1175 AspSerSerGluMetLysPheLysSerGluAsnGluAspTyrThrMetSerSerProLeuPro 1194
 |||||||
 Db 506 GATAGCTCTGAAATGATCATCAGTGAATGAAGATTATAC-ATGCTTCCACCTTTGGCA 564
 |||||||
 QY 1195 GlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPro 1214
 |||||||
 Db 565 GGAAGAAAAAGTGACAGAGAGAGAACGCTGATCTTGTAAAGTCTGAGCTGGAGAACCT 624
 |||||||
 QY 1215 ArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeu 1234
 |||||||
 Db 625 AGAAGTCGGAAGAAAAAGCAGCTGTCAAGACCTGAAAGAGAAATTAGTATGATGCTCA 684
 |||||||
 QY 1235 ThrLysLeuValGluGluGlnLysProLysGlySerGlnArgSerArg-LysArgGlyIle 1254
 |||||||
 Db 685 ACTAATGTGTGACAGACAGAAACCTAAAGCAGACAGCGCCGAAAAAGGCGCG 744
 |||||||
 QY 1254 sThraLaserGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIle 1274
 |||||||
 Db 745 TACAGCTTCACTCAGACGACAGCAGAGGCTGAGAGAGAGAGAGAGAGAGAGCT 804
 |||||||
 QY 1274 eLeuGluAsnGluAspGluGluAsnSerProProLysLysGlyLysArgGlyArgPro 1294
 |||||||

Db 805 CCTGGAATAATGAGATGACAGCAAAACAGCCCAAAAAAGGCAAAAGAGCAGGCCAC 864
 QY 1294 oLysPLeuGlyLysGlyLysThrProLysGluGluProThrMetLysTrpSerLys 1313
 |||||||
 Db 865 AAAACCTTGTGNTGGGGGACATCGAAGAGAGAAACCCACCATGAATCTCCAGAAA 923

RESULT 7
 CA324097
 LOCUS
 DEFINITION
 CA324097 786 bp mRNA linear EST 26-NOV-2002
 UT-M-FY0-cco-m-16-0-UT.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 IMAGE: 6822425 5', mRNA sequence.
 CA324097
 VERSION
 CA324097.1 GI:24542195
 EST.
 KEYWORDS
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 786)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lind, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

seq primer: pyx-5.
 Location/Qualifiers
 1..786
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6822425"
 /issue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_11b="NIH_BMAP_FY0"
 /note="Organ: Brain; Vector: pyx-Asc; Site_1: Ecor I;
 Site_2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 256 a 171 c 165 g 191 t 3 others

Alignment Scores:

Pred. No.: 6.18e-90 Length: 786
 Score: 1325.00 Matches: 256
 Percent Similarity: 97.71% Conservative: 0
 Best Local Similarity: 97.71% Mismatches: 6
 Query Match: 18.42% Indels: 0
 DB: 14 Gaps: 0

US-09-512-581b-2 (1-1391) x CA324097 (1-786)


```

Qy 932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProVal 951
Db 1 CGCTTACGGGCTCCCTGCTGATGACATGCGCATCTGTGCTCTTGGCCAAAGACCCCTGTG 60
Qy 952 LysGluTyrGArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
Db 61 AAAAGAGCGCAGCCATGATGACAGAGCTGTGGTGAAGAACATCATCACTGAGAGGGAG 120
Qy 972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuSerLeuLeuProGluTyrVal 991
Db 121 TACCTGAAGACGATGCGAGCTGTAGTCAAAAATATATGCTCTTCTTACCGAGATGTT 180
Qy 992 ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle 1011
Db 181 GTTCATATACAAATTCACCTTTGGACATGACCAGATATGTCTCAAGTACAGATAT 240
Qy 1012 GluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuGluIleLeuMetAlaLys 1031
Db 241 GAACAACTTAAGATGTAAGAAAGATGCTTTGGTTGGTGTGAGATATGATGCTTAA 300
Qy 1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLys 1051
Db 301 AATGAAAACACAGCATGATGATTTATCAGAAAATGTAGAAAATATTTAAACAGACAAA 360
Qy 1052 AspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspVal 1071
Db 361 GATGCTCAAGAGCAGATGATGATCAAAAATGATGATGATGATGATGATGATGATGAT 420
Qy 1072 AlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPro 1091
Db 421 GCCATGAACATCATCATCTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480
Qy 1092 ValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyr 1111
Db 481 GTGTGCGCAGCTCGGCTTTTCACCCAGCCTGACAAAGATTTAGTAAACACCAAAATATAC 540
Qy 1112 LeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeu 1131
Db 541 CTGCTCCAGAAAGAAATATCTTTCTCTCTGAAACCTGAAACCAAGCCAAATCTTCTC 600
Qy 1132 GlyAlaValAsnLysProLeuSerSerSerAlaGlyLysGlnSerGlnThrLysSerSerArg 1151
Db 601 GGAGCAGTATATAGCCACTTTCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 660
Qy 1152 MetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLys 1171
Db 661 ATGGAAGCTGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
Qy 1172 GlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSer 1191
Db 721 GGGAGGCTTGTATGCTCTGAAATGATCAGTGAAGATTAAGATTAACATATGCTTCA 780
Qy 1192 ProLeu 1193
Db 781 CCTTTG 786

```

RESULT 8
CA324284 779 bp mRNA linear EST 26-NOV-2002
LOCUS CA324284
DEFINITION UT-M-FY0-CCP-a-07-0-UI_r1 NIH_BMAP_FY0 Mus musculus cDNA clone
IMAGE: 6822512 5', mRNA sequence.
ACCESSION CA324284
VERSION CA324284.1 GI:24542382
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
Location/Qualifiers
1. 779
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6822512"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1lb="NIH_BMAP_FY0"
/note="Organ: Brain; Vector: pyx-Asc; Site:1: Ecor I;
Site:2: Not I. The library was constructed according
to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCAGACAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
Program coordinator."

BASE COUNT 241 a 166 c 174 g 197 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6,95e-89 Length: 779
Score: 1311.00 Matches: 255
Percent Similarity: 98.46% Conservatve: 0
Best Local Similarity: 98.46% Mismatches: 4
Query Match: 18.23% Indels: 1
DB: 14 Gaps: 0

US-09-512-581B-2 (1-1391) x CA324284 (1-779)

```

Qy 863 AspleuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAla 882
Db 3 GATTGACAGAAACAGGAAATTAAGTAACAGATATGACAGCGCTGAGACTGCTGCT 62
Qy 883 GlySerAlaIleValLysLeuAlaGlnLysProCysTyrHisGluIleIleThrLeuGlu 902
Db 63 GGGAGTGTATTGGAAAGCTGGACAGAGCCCTGTTACACGAGATCATCACTGGAG 122
Qy 903 GlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPhe 922
Db 123 CAGTACACAGCTGTGCTGATTTAGCCATCATATATAGTATGTTATCAAGTCAGCAGGTGTC 182
Qy 923 AlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIle 942
Db 183 GCTCAGAAACCTTCAACAAAGCGCTTCCCGCTTACGGCTTCCCTGAGTACATGGCCATC 242
Qy 943 CysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
Db 243 TGTGCTCTTTGTGCCAAAGACCCCTGTGAAAGAGAGCGAGCCCATGCTAGACACTGCTTG 302
Qy 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleValSerGluLys 982
Db 303 GTGAAGAACATCACTGTGAGGAGGAGTACTGTAAGCAGCATGAGCTGTTAGTGAATAA 362

```

QY 983 LeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAsp 1002
 |||||
 Db 363 TTAATTCTCTTTACACAGATATGTTGTCATATACATTCACCTTTGGACATGAC 422
 QY 1003 ProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1022
 |||||
 Db 423 CCAGATTATGTCAAACTACAGGATATTGAACAACCTTAAGATGTGAAGAATGCCCTTGG 482
 QY 1023 PheValLeuGluIleLeuMetAlaLysAsnGlnAsnSerHisAlaPheIleArgLys 1042
 |||||
 Db 483 TTTGTTCTGAGATATTGATGCTAAATAATGAATAACAGCCATCATTTATACAGAAA 542
 QY 1043 MetValGlnAsnIleLysGlnThrLysAspAlaGlnGlyProAspSpsAlaLysMetAsn 1062
 |||||
 Db 543 ATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGACCAATATACAAAATGAAT 602
 QY 1063 GluLysLeuTyrThrValCysAspValAlaMetAsnIleLeuMetSerLysSerThr 1082
 |||||
 Db 603 GAAAAATTGTACACCGCTGTGTGATGTTGCCATCAATCATCTATGCCAAGACACACAG 662
 QY 1083 TyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGlnProAsp 1102
 |||||
 Db 663 TACAGCCTGAGTCTCTAAGGACCCGCTGCTCCAGCTCGGTTTTCACCCAGCCTGAC 722
 QY 1103 LysAsnPheSerAspThrLysAsnTyrLeuProGluMetLysSerPhePheThr 1121
 |||||
 Db 723 AAGAAATTTAGTACACCANAAATTAACCTGCTCCAGAAATGAA-TCATTTTTCACCT 778
 RESULT 9
 BU703488 786 bp mRNA linear EST 09-OCT-2002
 LOCUS UI-M-FOO-bco-g-20-0-UI r1 NIH_BMAP_F00 Mus musculus cDNA clone
 DEFINITION IMAGE: 6405163 5', mRNA sequence.
 ACCESSION BU703488
 VERSION BU703488.1 GI:23629371
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS NIH-MGC (bases 1 to 786)
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabds-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 FEATURES
 source location/Qualifiers
 1..786
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6405163"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH-BMAP_F00"
 /note="Organ: Brain; Vector: pyx-Asc; Site1: Ecor I;
 Site2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TCAGAGAGCC. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 248 a 143 c 152 g 242 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6.16e-88 Length: 786
 Score: 1298.50 Matches: 254
 Percent Similarity: 98.08% Conservative: 1
 Best Local Similarity: 97.69% Mismatches: 4
 Query Match: 18.05% Indels: 1
 DB: 13 Gaps: 1

US-09-512-581b-2 (1-1391) x BU703488 (1-786)

QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGluValLys 20
 |||||
 Db 10 ATGGCTCATCAAGAA---AGACCACGATGGGAAAATTAATTAATTAATTAATTAATTA 66
 QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys 40
 |||||
 Db 67 GAAATCTCAGATTAATCTCTTAAGAGAGATGTCAGCGGTTAAAGATGCTGTGAAAA 126
 QY 41 ThrPheMetAspMetAspGlnAspSerGluGluGluMetValArgArgLeuLysMetVal 60
 |||||
 Db 127 ACTTATGATGACATGGAGCAGACCTGGAAGAGAAAGAAATTAATTAATTAATTAATTA 186
 QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
 |||||
 Db 187 TTACATCTTGCTTCTGCTGCTTCCCAAGACATCCGATAAAGATGTTGTTACCGTG 246
 QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
 |||||
 Db 247 GCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
 QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
 |||||
 Db 307 GATAAACTAAGGATATATTATGTTATTAACAGGCACTAAGGAGCTAGAGATACA 366
 QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
 |||||
 Db 367 AAGAGCCCTCAATTTAATAGATATTTTATTAATTAATTAATTAATTAATTAATTAATTA 426
 QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
 |||||
 Db 427 TATAACATATGCTTGTGAGTTAGAGATACCAATGAATCTTAACCTTAACAGAACAA 486
 QY 161 LeuPheSerValIleLeuAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeu 180
 |||||
 Db 487 TTAATCTCAGTTAATAACATGAGCCCAATCCAAAGATTCATATGACATGAGTGAACCTC 546
 QY 181 MetSerSerIleIleCysGluLysAspThrValSerGlnGluLeuLeuAspThrValLeu 200
 |||||
 Db 547 ATGAGCTATATCATTTGTGAGGTGATATTCATGAGACATTCGATGAGGTAAATCA 606
 QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrArgLeuAlaLysAla 220
 |||||
 Db 607 GTAATCTGTGTACCTGCTCCATGAAGAACTTAACCAAGCAATATGATTTGGCAAGCGT 666
 QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrPhePheAsnGlnVal 240
 |||||
 Db 667 TTGCTGAGAGAGACCTGTAACCTATGAAACCATATATTAACCAATTTTAAACAGGTT 726
 QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
 |||||

Db 727 CTCATGCTTGNGAACAATCATCATGATTGTCGTGAGCATGTCCTTGATTGATCTTG 786

RESULT 10
BU363024

LOCUS 603790485F1 CSEQCHN72 Gallus gallus cdna clone CHEST753121 5', mRNA
DEFINITION

ACCESSION BU363024 782 bp mRNA linear EST 28-NOV-2002
VERSION BU363024.1 GI:25871025
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

REFERENCE
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
1 (bases 1 to 782)
Phasianidae; Gallus.
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken CDNAS
Curr. Biol. 12 (22), 1965-1969 (2002)
JOURNAL 2235534
MEDLINE 12445392
PUBMED

COMMENT
Contact: Simon Hubbard
Department of Biomedical Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 10D, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..782
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST753121"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN72"
/note="Organ: Brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. CDNA
synthesis was initiated using an oligo(dt) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT 279 a 146 c 160 g 197 t

ORIGIN

Alignment Scores:
Pred. No.: 1.12e-87 Length: 782
Score: 1295.00 Matches: 253
Percent Similarity: 98.85% Conservative: 5
Best local Similarity: 96.93% Mismatches: 2
Query Match: 18.00% Indels: 2
DB: 13 Gaps: 0

US-09-512-581b-2 (1-1391) x BU363024 (1-782)

QY 856 ThrilleuHISerASpGIyASpIeuThrGIuGInGlyLysIleSerLysProAspMet 875
DB 3 ACAATATTACACAGTGTGAGACTTGACAGACGAGGAAATTTAGTAACCTGATATG 62

QY 876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysIleuAlaGInGluProCysTyr 895
DB 63 TCTCATGTGAGGTGGCGCGAGTGTGCTATTGTGAAGCTGGCACAAGAACTTGTGTAC 122

QY 896 HIsGluIleIeThrLeuGInGInTyrGInLeuGysAlaLeuAlaIleAsnAspGluCys 915
DB 123 CATGAATCATCACCCTTAGACAGTACAGCTGTGTGACTAGCCATAAACGATGAATGT 182

QY 916 TyrGInValArgGInValPheAlaGInLysLeuHIsLysGlyLeuSerArgLeuArgLeu 935
DB 183 TACCAAGTGAAGCAAAATATTGCTCAGAACTTCACAAGAGCGCTTCTAGACTACGGCTG 242

QY 936 ProLeuGluTyrMetAlaIleCysAlaLeuGysAlaLysAspProValLysGluArgArg 955
DB 243 CCACATGAAATATATGGCTATTGTGTGCACTGTGTCAAAAGATCCAGTCAAGAGAGA 302

QY 956 AlaHIsAlaArgGInGInLysValLysAsnIleAsnValArgArgGluTyrLeuLysGIn 975
DB 303 GCTCATGTGAGCGAGTGTGTGTGAAGACATAATGTCAAGAGGAATATCTGAACCAA 362

QY 976 HIsAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValAlaProTyrThr 995
DB 363 CATCAGACAGTTAGTGAAGAACTGTGTGCTTCTACAGAGTATGTGTTCCATTATACA 422

QY 996 ILeHIsLeuLeuAlaHIsAspProAspTyrValLysValGInAspIleGluGInLeuLys 1015
DB 423 ATACATCTTCTAGCTCATGAGCCAGATTATGTCAAGTCCAGACATGCAACCTAAG 482

QY 1016 AspValLysGluCysLeuTyrPheValLeuGluIleuMetAlaLysAsnGluAsn 1035
DB 483 GACATTAAAGAGTGTGTGTGTCTCATCTGGAATATGTAGTGTAAATAAGAGACACAC 542

QY 1036 SerHIsAlaPheIleArgLysMetValGluAsnIleLysGInThrLysAspAlaGInGly 1055
DB 543 AGTCATGATTTATACAGAAAATGTTGAATAATTTAAACAGACTAAAGATGCTCAAGGA 602

QY 1056 ProAspSpAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
DB 603 CCACATGATGCMAAAATGAATGAATAAACTCTACACAGCTGTGTGAGCAATGAATATC 662

QY 1076 ILeMetSerLysSerThrThr-TyrSerLeuGluSerProLysAspProValLeuProAl 1095
DB 663 ATTTATGTCAAGAGTACACCATACAGTTTGGAATCCCTTAACATCTCTGACTTCCAGC 722

QY 1095 aArgPhePheTherGInProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGlu 1115
DB 723 ACATATGTATCTCAACTGACAGAA-TTTAGTAACGCCAAAATATATCTTCTCCAAA 781

QY 1115 u 1115

DB 782 A 782

RESULT 11
BU914579 943 bp mRNA linear EST 17-OCT-2002
LOCUS AGNCCOURT_10496299 NICHD_XGC_001 Xenopus laevis cdna clone
DEFINITION IMAGE:6641106 5', mRNA sequence.
ACCESSION BU914579
VERSION BU914579.1 GI:24096493
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 943)
AUTHORS NCI-CCAP htcp://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Martha Rebber, Steven L. Klein, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLML14224 row: f column: 18
 High quality sequence stop: 746.

FEATURES

source

Location/Qualifiers

1..943

/organism="Xenopus laevis"

/mol_type="mRNA"

/db_xref="taxon:8355"

/clone="IMAGE:6641106"

/tissue_type="oocytes"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NICHD_XGC_001"

/note="Vector: PCMV-SPORT6; Site.1: NotI; Site.2: SalI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 2.2 kb. Constructed by Life Technologies."

BASE COUNT

297 a 182 c 215 g 249 t

ORIGIN

Alignment Scores:

Pred. No.:	7.38e-87	Length:	943
Score:	1285.50	Matches:	261
Percent Similarity:	87.38%	Conservative:	16
Best Local Similarity:	82.33%	Mismatches:	30
Query Match:	17.87%	Indels:	10
		Gaps:	5

US-09-512-581b-2 (1-1391) x BU914579 (1-943)

QY 744 PheaaGlnIlePheGluProLeuHISySerLeuAspProSerAsnLeuGlnHISLeu 763
 |||||
 Db 1 TTGGACAGATATTGAGCCCTTACATAAAGCTTGACCCCTGGCAACCTGAGCACTG 60
 |||||
 QY 764 IleHrProLeuValThrIleGlnHISleuLeuAlaLeuAlaProAspGlnPheAla 783
 |||||
 Db 61 ATCACATCAGTCGTCATGGCCACATAGCACAGCTTGTCCAGATCAGTCACTGCT 120
 |||||
 QY 784 ProTrpLysSerTrpValAlaThrPheIleValAspPLeuLeuMetAsnAspArgLeu 803
 |||||
 Db 121 CCTTAAATCATCAGGTGGCCACCTTGTGTTAAAGTTTGTATGATCAGATGGCTT 180
 |||||
 QY 804 ProGlnLysLysThrThrLysLeuTrpValProAspGlnGlnValSerProGlnThrMet 823
 |||||
 Db 181 CCAGGTAAAGAACAAACAACTCGGCTGCATGATGAGGTGCGACAGAAACTAAG 240
 |||||
 QY 824 ValLysIleGlnAlaIleLeuMetMetValArgTrpLeuLeuGlnMetLysAsnHIS 843
 |||||
 Db 241 GTTAAATTCAGCAATTAATGATGGTGGGCTGGCTGGGGAATGAAAATATATCTC 300
 |||||
 QY 844 SerLysSerGlnThrSerThrLeuArgLeuLeuThrThrIleLeuHISerAspLysAsp 863
 |||||
 Db 301 AGTAAATCCGGGAATTCACCTTACGACTCTTAATGGAATATTTCCACACATCAGGAGAC 360
 |||||
 QY 864 LeuThrGlnGlnGlnLysIleSerLysProAspMetSerArgLeuArgLeuAlaIleGly 883
 |||||
 Db 361 TTGACAGAACATCGAAAGCTCAGTAAACCAACATGTCACGCTGAGACTGCTCTGGC 420
 |||||
 QY 884 SerAlaIleValLysLeuAlaGlnGlnProCysTrpHISGlnIleIleThrLeuGln 903
 |||||
 Db 421 AGTGCATTTGTTAAATGAGCGACGACGATGTTACATGAAATTAATTAACCTTAACACAG 480
 |||||
 QY 904 TyrGlnLeuCysAlaLeuAlaIleAsnAspGlnCysTrpArgGlnValArgGlnValPheAla 923
 |||||
 Db 481 TATCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 |||||
 QY 924 GlnLysLeuHISLysGlnLysSerArgLeuArgLeuProLeuGlnLysMetAlaIleCys 943
 |||||
 Db 541 CAAAAAATTCACAAAGGCTCTCCAGATTAAAGACTTCATTAATATATGCGCATCTGCT 600
 |||||

QY 944 AlaLeuCysAlaLysAspProValLysGlnArgArgAlaHISAlaArgGlnCysLeuVal 963
 |||||
 Db 601 GCTGTGTGCAAAAGATCTGTGAAGAAGACGCGTCACGACGAGTGTGTG 660
 |||||
 QY 964 LysAsnIleAsnValArgGlnLysTrpLeuLysGlnHISAlaAlaValSerGlnLysLeu 983
 |||||
 Db 661 AAAAATATTATATGTCAGAAAGAGATATGTAAGACAGATGCGCTGAGTAAATAATG 720
 |||||
 QY 984 LeuSerLeuLeuProGlnTrpValValProTrpThrIleHISLeuLeuAlaHISAspPro 1003
 |||||
 Db 721 TTCTCTCTTCGCTGCAATATGCTGCTCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 |||||
 QY 1004 AspTrpValLysValGlnAspTrpLeuLeuLeuLysAspValLysGlnCysLeuTrp 1022
 |||||
 Db 781 GATTTATGTGGAAGTCAAGATATAGAGCAGCTAAAGGACATCAAGAGAG 834
 |||||
 QY 1023 -----PheValIle----GlnIleLeuMetAlaLysAsnGlnAsnSerHISAlaPhe 1039
 |||||
 Db 835 GTGGGCTTCGTCGCTGCGAGATTTCTTATGTCAAGAAATGAAATTAACAGCATGCTT 894
 |||||
 QY 1039 e---IleArgLysMet-----ValGlnAsnIleLysGlnThrLysAsp 1052
 |||||
 Db 895 TTAAATTCGAAAAAATGGGTAGAAATATTAATTAAGCAAAACAAAGAA 943
 |||||

RESULT 12

CA987660

LOCUS CA987660 857 bp mRNA linear EST 06-JAN-2003

DEFINITION

AGENCOURT_11108371 NICHD_XGC_Emb1 Xenopus laevis cDNA clone

ACCESSION

IMAGE:6864353 5', mRNA sequence.

VERSION

CA987660.1 GI:27520335

KEYWORDS

EST.

SOURCE

Xenopus laevis (African clawed frog)

ORGANISM

Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE

1 (bases 1 to 857)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgap@remail.nih.gov

Tissue Procurement

Martha Rebert, Steven L. Klein, Ph.D.

cDNA Library Preparation

Life Technologies, Inc.

DNA Sequencing

Agencourt Bioscience Corporation

Clone distribution

NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLML14484 row: p column: 16
 High quality sequence stop: 670.

FEATURES

Location/Qualifiers

1..857

/organism="Xenopus laevis"

/mol_type="mRNA"

/db_xref="taxon:8355"

/clone="IMAGE:6864353"

/tissue_type="embryo (stage 10)"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NICHD_XGC_Emb1"

/note="Vector: PCMV-SPORT6; Site.1: NotI; Site.2: SalI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 1.55 kb. Constructed by Life Technologies. Note: This
 is a Xenopus Gene Collection (XGC) library."

BASE COUNT

275 a 160 c 192 g 230 t

ORIGIN

Alignment Scores:

1.06e-85 Length: 857

Score:

1269.50 Matches: 249

Percent Similarity:

91.81% Conservative: 9

Best Local Similarity:	88.61%	Mismatches:	21
Query Match:	17.65%	Indels:	2
DB:	14	Gaps:	1
US-09-512-581B-2 (1-1391) x CA987660 (1-857)			
OY	784	ProTrpLysSerThrValAlaIleThrPheIleValLysAspLeuMetAsnAspArgLeu	803
DB	3	CCCTTAAATATCCAGTGGTGGCCACCTTGTGGTTAAAGATTGTGTATGATACAGATGGCGTT	62
OY	804	ProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMet	823
DB	63	CCAGGTAAAAAGACAAACAAACCTCGGGTGCACATGATGAGGTGTCCAGCAAACTAG	122
OY	824	ValLysIleGlnAlaIleLysMetMetValArgTrpLeuGluGlyMetLysAsnHis	843
DB	123	GTTAAATATTCAGACATTAATAATATGATGGTGGGAGGCTGGCTGGGATGAAAAATATCTC	182
OY	844	SerLysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHisSerAspGlyAsp	863
DB	183	AGTAAATCCGGGAATTTACTTTACGACTCTTAATGGCAATATTCACACATGACGGAGAC	242
OY	864	LeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaIleGly	883
DB	243	TTGACAGAACATGGAAAGCTCAGTAAACCAACATGACGCTGACGCTGACTGCTCTCC	302
OY	884	SerAlaIleValLysLeuAlaGlnGluProCysTryHisGluIleIleThrLeuGluGln	903
DB	303	AGTGCATATGTTAAACTGACGCGCAGGACATGTTACATGAAATATTACTTGAACAG	362
OY	904	TyrGlnLeuCysAlaLeuAlaIleAsnAspGlyCysTyrGlnValArgGlnValPheAla	923
DB	363	TATCAGCATATGTGCTTGGTGCATCATGATAGTGTACCGAGGTAGACAGTGTCTTGGCT	422
OY	924	GlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys	943
DB	423	CAAAAAATTCCAAAGGGCTCTCCGATTAAGACTTCCATTGAAATATATGCGCATCTGT	482
OY	944	AlaLeuCysAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuVal	963
DB	483	GCTCATATGCAAAAGATCCCTGTGMAAGAGAGCGCCGCTCAGCGCAGAGTGTGGTG	542
OY	964	LysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeu	983
DB	543	AAAAATATTAATGTCAGAGAGAGATGATGTGAACAGCATGACGCTGAGTGAATAAATTG	602
OY	984	LeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspPro	1003
DB	603	TTTCTCTTTCGCGCGTAATATGTGGTTCATATACGTGTTCAATTGCTTGCAATGACCCA	662
OY	1004	AspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpH	1022
DB	663	GATTATGGAAGATGACAGATATGAGCAGCTAAAGACATCAAAAGATGGCGCTGTGTT	722
OY	1023	eValLeuGluLysLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMe	1043
DB	723	CGTGTGGAATTTCTTATGTCGAAGAATGAATAATACAGCCATGCTTTATTTCCGAAT	782
OY	1043	IvaGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnG	1063
DB	783	GGTGAATATATTTAAGCCACCAAAAGATGGCCCAAAATCTGATGACCAAG--AATGA	839
OY	1063	u 1063	
DB	840	A 840	
RESULT 13			
BU916668	882 bp	mRNA	linear
LOCUS	AGNC00008.10496784	NICHD_XCC_CO1	Xenopus laevis cDNA clone
DEFINITION	IMAGE:0643610 5', mRNA sequence.		
ACCESSION	BU916668		
VERSION	BU916668.1	GI:24098582	

KEYWORDS	EST.
SOURCE	Xenopus laevis (African clawed frog)
ORGANISM	Xenopus laevis Eumetazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus. 1 (bases 1 to 882)
REFERENCE	NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap .
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE	Unpublished
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgaaps@email.nih.gov
COMMENT	Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM14230 row: 0 column: 02 High quality sequence stop: 663. Location/Qualifiers 1..882
FEATURES	/organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:66310" /clone="IMAGE:6643610" /issue_type="ocytes" /_lab_host="DH10B (phage-resistant)" /_clone_id="NICHD_XC_001" /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Life Technologies."
BASE COUNT	310 a 169 c 192 g 210 t 1 others
ORIGIN	
Alignment Scores:	
Pred. No.:	3,69e-85 Length: 882
Score:	1262.50 Matches: 246
Percent Similarity:	91.35% Conservative: 18
Best Local Similarity:	85.12% Mismatches: 25
Query Match:	17.55% Indels: 1
DB:	13 Gaps: 0
US-09-512-581B-2 (1-1391) x BU916668 (1-882)	
Oy	556 AAlHIsAlaATgGInCysLLeuValLyAsnLIleAsnValATgATgAGLUtTyLeuLySGln 975
Dd	3 GCCTACGCACGGCAGCTGTCTGTGTAATAATTATATGCAAGAAGACTATCTGAACAG 62
Oy	976 HIsAlaAlaValSerGIuLySLeuLeuSerLeuLeuProGIuTyValValProTYrThr 995
Dd	63 CATGCAGCTGTGAGTAGAAAAATTTGTCCTCTCTGCTGAGTAATGTGGTTCATATACT 122
Oy	996 ILlHtsILeuAlaHtsASpproasprryrValLySVAlgluAspIlleGlInleuLyS 1015
Dd	123 GTTATTTGGTGTGCATGACCACCGATTATGTGAAGACAAAGATATAGACACTAAAG 182
Oy	1016 ASpValLySGluCySLeuTrPrPhVaLIleuLIuLIleuMeTaLaLySAsnGLuaSnAsn 1035
Dd	183 GACATCAAAGAGTGCTGTGTGTGTGTGTGAATTTATGTCAAAGAAATGAAMAATAC 242
Oy	1036 SerHisAlaPheIIeaTgLySMetValGlUaSnIlleLySGInThrLySASPalaGlInCly 1055
Dd	243 AGCATGTGTTTTATTGCGAAAAATGTGTGAATATTTTAAAGCAAAAAGATGCCCAAAC 302
Oy	1056 ProkAsPaSPalALySMetAsnGLuLySeuTyrrThrValLySAsPvalAlAMetAsnIlle 1075
Dd	303 CCCCATGACCGAAGATGAATGAATAAATGTACACAGTGTGTGATGTGGCATGAATATA 362
Oy	1076 IlleMetSerLySerThrThrTyLSerLeuGlueSerPolySAsPProValLeuProAla 1095

Db	Accession	Version	Keywords	Source	Organism	Reference Authors Title	Journal Comment	Features Source
Db	363	ATTATTTTCAAGACGACACACACTACAGCTTAAGATGCCGGAAGGATCCTGATTCGCTGCA	422					
Qy	1096	ArgpPhePheThgInPProAspLysAsnPhSeSerAsnThrLysAsnTyLLeuProProGlu	1115					
Db	423	CGTTTCTTCATCTACGACGAGATTAAGAAATTTTCAGCAACACAAAAAATTATCTCCCGCTGAA	482					
Qy	1116	MetLysSerPhePheThgProGlyLysProLysThrThrAsnValLeuGlyAlaValAsn	1135					
Db	483	TTGGAAGTCCTCTTTACACCGAGGAAGCCAAAATCTACTAATGTCTTGCGTCACTCAAT	542					
Qy	1136	LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal	1155					
Db	543	AAACCTTATGACGACGAGGAAGAACAAATGCTATCCAAAGTATCTAGATGACACTGTA	602					
Qy	1156	SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgGlyLeuGlyArgLeuAsp	1175					
Db	603	AGCAATGCAAGACGCGGTTTCAACCCAAAGCTCTCCAGGCAAGATTAAGAAAGAACTTAC	662					
Qy	1176	SerSerGluMetAspHisSerSerGluAsnGluAspLysPylrThrMetSerSerProLeuProGly	1195					
Db	663	AGTGTGGAATGTGATCAAGAGTAAAGTAAACATATACGATGTCCTTCCTCTCTGCGG	722					
Qy	1196	LysLysSerAspLysArgAspAspSerSerAspLeuAlaGlySerGluLeuGluLysProArg	1215					
Db	723	AAAAAAGTGACAAAAGGAGCAGATCTCGATCTCTTGAAGTCTGACTTGAAAAGCCAAAG	782					
Qy	1216	GlyArgLysLysThrProValThrGluGlnGluLysLysLeuGlyMetAspAspLeuThr	1235					
Db	783	CG-AGGGAAGAACAATCTCTTATACCAACATGACAGTCTTAAGCATGATGAATATATC	841					
Qy	1236	LysLeuValGlnGlnGluLysProLys	1244					
Db	842	AAACCCAGCACAGGGAACCAAAATCAG	868					
RESULT 14								
LOCUS	BU914025	931 bp	mRNA	linear	EST 17-OCT-2002			
DEFINITION	AGNCNCOURT_104944702 NICHD_XGC_001 Xenopus laevis cdna clone							
ACCESSION	IMAGE:6640480 5', mRNA sequence.							
VERSION	BU914025							
KEYWORDS	BU914025.1 GI:24095939							
SOURCE	EST.							
ORGANISM	Xenopus laevis (African clawed frog)							
	Xenopus laevis							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;							
	Xenopodidae; Xenopus.							
REFERENCE	1 (bases 1 to 931)							
AUTHORS	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.							
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),							
	Tumor Gene Index							
	Unpublished							
JOURNAL	Contact: Robert Strausberg, Ph.D.							
COMMENT	Email: cgapbs-remail.nih.gov							
	Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.							
	CDNA Library Preparation: Life Technologies, Inc.							
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)							
	DNA Sequencing by: Agencourt Bioscience Corporation							
	Clone distribution: NCI-CCAP clone distribution information can be							
	found through the I.M.A.G.E. Consortium/LNLN at:							
	http://image.lnl.gov							
	Plate: LLM14222 row: 1 column: 16							
	High quality sequence stop: 552.							
	Location/Qualifiers							
	1..931							
	/organism="Xenopus laevis"							
	/mol_type="mRNA"							
	/db_xref="taxon:8355"							
	/clone="IMAGE:6640480"							
	/tissue_type="Oocytes"							
	/lab_host="DH10B (phage-resistant)"							

BASE COUNT	286 a	173 c	202 g	269 t	1 others
ORIGIN	Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Life Technologies."				

Alignment Scores:

Pred. No.:	2,33e-83	length:	931
Score:	1299.00	Matches:	250
Percent Similarity:	93.88%	Conservative:	8
Best Local Similarity:	91.01%	Mismatches:	14
Query Match:	17,23%	Indels:	4
DB:	13	Gaps:	0

US-09-512-581B-2 (1-1391) x BU914025 (1-931)

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: September 25, 2003, 00:10:16 : Search time 852 Seconds

(without alignments)
4064.415 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKRTNDGRTYPGVK.....QKGRGRPSKTPSPQPKKNV 1391

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1678620 seqs, 1244745471 residues
Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163549_5590/app.query.fasta_1.1543
-DB=Published_Applications_NA -OFMT=fastlap -SUFFIX=rrpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MITLEN=0
-MAXLEN=200000000 -USER=US09512581.ecgn_1.1.474_errunat_23092003_163549_5590
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEOUTPUT -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-GAPOP=6 -GAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1409.5	19.6	2496	9	US-09-822-849A-496 Sequence 496, App

	2	762.5	10.6	1217	9	US-09-925-302-357	Sequence 357, App
	3	698	9.7	439	13	US-10-040-739-107	Sequence 107, App
	4	598	8.3	471	11	US-09-918-995-3514	Sequence 3514, App
	5	349	4.9	530	10	US-09-879-536-682	Sequence 682, App
	6	276	3.8	433	11	US-09-918-995-20720	Sequence 20720, A
	7	276	3.8	433	11	US-09-918-995-33817	Sequence 33817, A
	8	232	3.2	8493	13	US-10-071-766-51	Sequence 51, Appl
	9	231.5	3.2	9274	10	US-09-885-535-3	Sequence 3, Appl
	10	221	3.2	8503	14	US-10-037-270-130	Sequence 130, App
	11	229	3.2	7992	11	US-09-893-519A-140	Sequence 140, App
	12	229	3.2	8491	12	US-10-133-013-260	Sequence 260, App
	13	228	3.2	7034	14	US-10-118-513A-11	Sequence 11, Appl
	14	225	3.1	5334	12	US-09-882-227-623	Sequence 623, App
	15	205.5	2.9	3727	14	US-10-171-581-49	Sequence 49, Appl
	16	205.5	2.9	7659	12	US-10-096-534-34	Sequence 34, Appl
	17	205.5	2.9	8063	12	US-09-814-353-21776	Sequence 21776, A
	18	205	2.8	7497	10	US-09-960-253-175	Sequence 175, App
	19	205	2.8	7792	13	US-10-044-090-359	Sequence 359, App
	20	200.5	2.8	4883	14	US-10-118-513A-5	Sequence 5, Appl
	21	200.5	2.8	7215	14	US-10-118-513A-13	Sequence 13, Appl
	22	200.5	2.8	7215	14	US-10-118-513A-15	Sequence 15, Appl
	23	200.5	2.8	10281	14	US-10-084-817-361	Sequence 361, App
	24	200	2.8	10281	12	US-10-240-965-201	Sequence 201, App
	25	197.5	2.7	10300	10	US-09-960-253-145	Sequence 145, App
	26	197	2.7	10190	10	US-09-864-864-292	Sequence 292, App
	27	197	2.7	10211	10	US-09-954-456-1153	Sequence 1153, App
	28	197	2.7	10211	10	US-09-967-768A-186	Sequence 186, App
	29	197	2.7	10211	11	US-09-918-624B-2	Sequence 2, Appl
	30	196	2.7	10096	10	US-09-960-253-163	Sequence 163, App
	31	194	2.7	5967	14	US-10-171-581-158	Sequence 158, App
	32	193	2.7	5612	14	US-10-128-714-1388	Sequence 1388, App
	33	193	2.7	5612	14	US-10-128-714-6388	Sequence 6388, App
	34	193	2.7	7611	14	US-10-128-714-388	Sequence 388, App
	35	193	2.7	7612	14	US-10-128-714-3588	Sequence 3588, App
	36	191	2.7	4455	11	US-09-998-027-7	Sequence 7, Appl
	37	191	2.7	10302	10	US-09-782-378A-23	Sequence 23, Appl
	38	191	2.7	16531	12	US-10-101-510-667	Sequence 667, App
	39	190.5	2.6	2212	9	US-09-925-297-142	Sequence 142, App
	40	189.5	2.6	1639	12	US-10-017-161-2225	Sequence 2225, App
	41	189	2.6	5397	14	US-10-128-714-7388	Sequence 7388, App
	42	189	2.6	12337	14	US-10-171-311-5	Sequence 5, Appl
	43	189	2.6	12462	14	US-10-171-311-1	Sequence 1, Appl
	44	187.5	2.6	6452	14	US-10-205-823-418	Sequence 418, App
	45	186.5	2.6	5194	11	US-09-998-027-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-822-849A-496
Sequence 496, Application US/09822849A
Patent No. US20020045170A1
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulikota, Kamalakkar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6403
CURRENT APPLICATION NUMBER: US/09/822,849A
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/195,582
NUMBER OF SEQ ID NOS: 598
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 496
LENGTH: 2496
TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-822-849A-496

Alignment Scores:
Pred. No.:      8,26e-118      Length:      2496
Score:          1409.50       Matches:      300
Percent Similarity: 65.96%    Conservative: 76
Best Local Similarity: 52.63%  Mismatches:   123
Query Match:     19.60%      Indels:       72
DB:              9           Gaps:          8

US-09-512-581b-2 (1-1391) x US-09-822-849A-496 (1-2496)

QY 717 LysSerLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
    |||:||||:||||| ||| ||||| |||:||||| ||||| |||
Db 4 AAGACAAAGAGGGGTACTCCACACACGAAACAAAGCGCTGCCTATACAGCCATA 63

QY 737 PheSerSerLysGluThrGlnPheAlaGlnIlePheGluProIleHisLysSerLeuAsp 756
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 64 TTCACAAATTAAGAAAGTCACAGCTGCACAGATTTTTCAGTCA----- 105

QY 757 ProSerAsnLeuGluHisLeuIleThrProLeuValIleThrIleGlnHisIleAlaLeuLeu 776
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 105 ----- 105

QY 777 AlaProAspGlnPheAlaAlaProThrLysSerSerTrpValAlaThrPheIleValLysAsp 796
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 105 ----- 105

QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 106 -----ACAGGTGAAAAGAAATGCAAAAGCTGGTCTCCAGATGAA 144

QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 145 GAGGTTCCCTGAAAGTACTAGCAAGAGTTCAGGCATTTAACTTCGGTAGAGTGGCTG 204

QY 837 LeuGluMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThr 856
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 205 TTGGGTATGAAAAACAACAGCTGAATCTGCACATCAACCCCTGGTATTATTCAGCGC 264

QY 857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 265 ATGTGGTATGAGGGGTGAGCTGACAGCAACAAAGAGATCACTGAATATCT 324

QY 877 ArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 325 CGCTGGGATTAGCTGCTGCTAGTGCATATATGAAGCTTGCTCAGAACTTGTTACAT 384

QY 897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 385 GAATATTATTCACCCAGAAAGAGTTTCAGCTGTGCTGCTGTTATTATGATGAGTGTAC 444

QY 917 GluValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerAlaGlyLeuArgPro 936
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 445 CAAGTAAAGGCAATATTGCTCAGAACTGCATTAAGCCACTTGTAAGATTACTGCTCCA 504

QY 937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 505 TTGGAGATATATGGCGATCTTGCTTGCTGTGGCCAAAGATCTGTGAAAGGAGAGAAAGCA 564

QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 565 CACGACAGCAATAGTTACTGAAATAATATACGTATACGACGAGGAATACATTAAACACAAT 624

QY 977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 625 CCTATGGCTACTGAAATATATATATACCTGTGGCTTAATAATAGTAGTTCATACATAT 684

QY 997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 685 CACCTGCTAGCCCATGATCCAGATTTTACAAAGATCACAAAGATGTGTATGACGCTTGAT 744
```

```
QY 1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
    ::|||:||||:||||| ||| ||||| ||| ||||| |||
Db 745 ATCAAAAGAGTCCATATGTTCTGATGCTTGAAGTTTATGACAAAGATGAAACAAATAGC 804

QY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 805 CATGCCCTTATGAAAGAGATGGCGAGAACATCAAGTTAAACAGAGATGCCAGCTTCYCA 864

QY 1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 865 GATCAATCCAGAACAAATGAAAATGTATACATGATATGATGCTGCTCTGTGTATATA 924

QY 1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 925 AATAGTAAAGTGGCTTGT--GCAATGCAGATTTCCAAAGAGACCACCTCCAAATGAAT 982

QY 1097 PheThrThrGlnProAspLysAsnPhSerAsnThrLysAsnTyrLeuProProGluMet 1116
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 983 TTTTACACA---CCTGAAAAGGACTCTGTAAACCATGAAGATTATATTTCAGAAAGACAA 1039

QY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGluAlaValAsnLys 1136
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1040 AGACTACTTCTGTTAACAGGAAAGCCAAAGCCCTCTGAGTACTAGTGCAGTAAATTAAG 1099

QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1100 CCTTATCAGACAGGAGAGAAACCTATGTTAGAGCACACTGCACCTGAGACTGCAAGC 1159

QY 1157 -----AsnAlaSerSerSerSerAsnProSerSerProGluArgIleLysGlyArg--- 1173
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1160 AATATTATGTAAATTGATGAGCTGAAACCTTCAACCGGAAATGCATCAAGGAAACAGAT 1219

QY 1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp-----TyrThrMetSerSer 1191
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1220 TCAGAGCGACAGAGAACTGGAGTATGTAAGAAATTAACAGAAACCTCGAGAGATTATTTCA 1279

QY 1192 ProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeu 1211
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1280 GTCACACCTGTAAGATATGATGACCCAGTAAAGAAATGAG-----GTAAGAAATGCATTTG 1333

QY 1212 GluLysProArgLysArgLysLysThrProValThrGluGlnGlu-GluLysLeuGlyMe 1231
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1334 CAAGGGAGAAATGAAGGCCAAACAGAACAGAGGCTCAGCTTGCAGAAAACCTTGATG 1393

QY 1231 LAspAspLeu-----ThrLysLeuValGlnGluGln-----Lys 1242
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1394 CACAAATGTCCTCTGAACAGAAATGAAGCTCAGTTCAGAACACACACTCTGCTTGAA 1453

QY 1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1262
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1454 AACTTAAGAGACATGTTACTTCTCTTTCACATGACCAAGTCCCTGATGGAATGTACA 1513

QY 1262 ngIntPrProGluGluLysArgLeuLys 1271
    |||:||||:||||| ||| ||||| ||| ||||| |||
Db 1514 GCAGAAACTCTTGAGAGAGAGGCTAATAA 1541

RESULT 2
US-09-925-302-357
; Sequence 357, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
```

```

; LENGTH: 1217
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1141)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1157)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-357

```

```

Alignment Scores:
Pred. No.: 1.76e-59 Length: 1217
Score: 762.50 Matches: 185
Percent Similarity: 56.88% Conservative: 63
Best Local Similarity: 42.43% Mismatches: 121
Query Match: 10.60% Indels: 68
Gaps: 15

```

US-09-512-581b-2 (1-1391) x US-09-925-302-357 (1-1217)

```

OY 965 AsnIleAsnValArgGluTyrLeuLysGlnHisAlaIaValSerGluLysLeu 984
    |||.....|
DB 2 AATATCAATATACGAGGGAATACATTAAAGCAGATCTATGCTACTGAGAAATTATTA 61
OY 985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
    |||.....|
DB 62 TCACGTGGCTGGAATATGATGTTCCATACATGATTCACCTGATGCCCATGATCCAGAT 121
OY 1005 TyrValLysValGlnAspIleGlnIleuLysAspValLysGluCysLeuThrPheVal 1024
    |||.....|
DB 122 TTACACAGATCAACAAGATGTGATGCTCTGCTGATATCAAGAGCTGCTATGATGTTGATG 181
OY 1025 LeuGluIleLeuMetAlaLysGlnLysAsnSerHisAlaPheIleArgLysMetVal 1044
    |||.....|
DB 182 CTGGAAGTTTAAATGACAAAGATGAAACATATACCATGCTTTATGAGAAAGTGGCA 241
OY 1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnLys 1064
    |||.....|
DB 242 GAGAACATCAAGTAAACAGAGATGCCAGTCCAGTCCAGATGATCCAGCAATGATAA 301
OY 1065 LeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSerThrTyrSer 1084
    |||.....|
DB 302 CTGATATACAGTATGATGTGCTCTGCTGTTATTAATAGTAAAGTCTTGTGCAAT 361
OY 1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
    |||.....|
DB 362 GCAGATTACCAAGAGACCCAGCTCTCCCATGAAATTTTACCAACCACTGAAAGGAC 421
OY 1105 PheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrProGlyLys 1124
    |||.....|
DB 422 TTCTGTAACGATTAAGAGTTATATTACAGAGACAGACAGATGACTCTGTTAAACAGGAAG 481
OY 1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGln 1144
    |||.....|
DB 482 CCAAGCCTGCTGAGTACTAGTGCAGTCAATAAAGCCTTATACAGCAACGGGAAGAAA 541
OY 1145 SerGlnThrLysSerSerArgMetGluThrValSer-----AsnAlaSerSerSer 1162
    |||.....|
DB 542 CCCATATGTTAAGACAGACGCGACAGTGCAGACATATTAATGTAATTCACAGCTG 601
OY 1163 AsnProSerSerProGlyArgIleLysGlyArg--LeuAspSerSerGluMetAspHis 1181
    |||.....|
DB 602 AACCTTCACCGGAAATCGATCAAGGAGCAAGTTCAGAGCGCAGAAACTGAGACT 661
OY 1182 SerGluAsnGluAsp-----TyrThrMetSerSerProLeuProGlyLysLysSerAsp 1199
    |||.....|
DB 662 AGTGAATAATGAAGAACCCCTGTGTGAGGATTTATTTCAGTCACACCTGTAAGAATATTGAC 721

```

```

OY 1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys 1219
DB 721 -----
OY 1220 ThrProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGln 1239
    |||.....|
DB 722 ---CCAGTAAAGAAATAAGAAA-----ATTAAATCTGTATCAGGCTAC----- 760
OY 1240 GluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
    |||.....|
DB 761 ---CAGGCAACATCAGCAGTCCGCGCCCTTCCAAACCCAGAGAGACGTGACCCAAAGCTGAA 817
OY 1260 AspGlu---GlnGlnThrProGluGlnLysArgLysGluAspIleLeuGlnAsnGlu 1278
    |||.....|
DB 818 GAGATATCCAAACAAAAACAGATGAGAAA-----GTA 850
OY 1279 AspGluLysAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeu 1297
    |||.....|
DB 851 GATGAATCGGCGACCTCCGCGCCCTTCCAAACCCAGAGAGACGTGACCCAAAGCTGAA 910
OY 1298 Gly---GlyGlyThrProLysGluGluLysProThrMetLysThrSerLysGlySerLys 1316
    |||.....|
DB 911 TCTCAGGCAATGCTACCAAAAATGATGATCTAAATTAACCTATTAAACAGGAGGAAG 970
OY 1317 LysLysSer---GlyProProAlaProGluGluGlu----- 1328
    |||.....|
DB 971 AGAGCTGCAGTGGCTCAGAGAGACCTCGGGGTTTGGAAAGCAAGTATGCCAAAGACACC 1030
OY 1339 -----GluGluGluArgGlnSerGlyAsnThrGlnLys-----SerLys 1342
    |||.....|
DB 1031 AAACGCAAGATTT-AGCCAAAAGCAGACACAGACAGAAAGACMAATTGACTTACAAAG 1089
OY 1343 SerLysGlnHis-----ArgValSerArgArgAlaGlnGlnArgAlaGluSerProLys 1360
    |||.....|
DB 1090 GTRAAATGCTTTTGCAAAAGGAGAAATGAAGGCCAAACAGAACCCAGCTNCCAGYTTN 1149
OY 1361 SerSerAlaIle-----GluSerThrGlnSerThr 1370
    |||.....|
DB 1150 TGCAAAANCTTGATTAACATGKCTGACAGAAATGACTTATTCACACC 1197

RESULT 3
US-10-040-739-107
; Sequence 107, Application US/10040739
; Publication No. US20020173635A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John
Lavallee, Edward
Racie, Lisa
Merberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998

```



```
; ORGANISM: Homo sapiens
US-09-879-536-682

Alignment Scores:
Pred. No.: 1.87e-22 Length: 530
Score: 349.00 Matches: 67
Percent Similarity: 98.53% Conservative: 0
Best Local Similarity: 98.53% Mismatches: 1
Query Match: 4.85% Indels: 0
DB: 10 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-879-536-682 (1-530)
QY 37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluLysGluLysLeuTyr 56
DB 213 ATGGTTGTAAGACTTTATGATATGAGACGAGCTCTGAGAGAAAGAGAGCTTTAT 154
QY 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
DB 153 TTAACCTAGCTTTACATCTTGCTTCAGATTTTCCCAAGCATCTGATATAGATGTT 94
QY 77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
DB 93 CGCTTACTGTAGAGCTGCTGCTGCTGATATTTTCAGAGATTATGCTCTGAAAGCTCCT 34
QY 97 TyrThrSerProAspLysLeuLys 104
DB 33 TACACATCCCTGATTAACCTAAG 10

RESULT 6
US-09-918-995-20720
; Sequence 20720, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20720
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-20720

Alignment Scores:
Pred. No.: 5.49e-16 Length: 413
Score: 276.00 Matches: 62
Percent Similarity: 52.94% Conservative: 10
Best Local Similarity: 45.59% Mismatches: 7
Query Match: 3.84% Indels: 57
DB: 11 Gaps: 1

US-09-512-581b-2 (1-1391) x US-09-918-995-20720 (1-413)
QY 11 GlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluLys 30
DB 7 GGCACGAGCGCTTACCTCCGCGGGGTAAAGATACCGACAGATCAACCAAGAGAGAG 66
QY 31 MetValArgArgLeuLys----- 36
DB 67 ATGATCAAAAGCCGTGAAGAGGAGAAACCAACTGCTTTTCACTTGAGCTGTCTAGAAAGC 126
QY 36 ----- 36
DB 127 CCATGTAGTGTGAGACTGAAGCCTTCAACAACAGCAAGTGTCAACTTATAACCGTAGA 186
QY 36 ----- 36
```

```
DB 187 ATGACTTACCTTCAAAACACACACCCCTCCAGCCCACTCAAGCCTTCAGATTAAGTCAAGCCT 246
QY 37 -----MetValValLysThrPheMetAspMetAspGlnAspSerGluGluLysGlu 54
DB 247 TGACACAGATGTAGTGAAGAACTTTATGATATGATGATCAGAGCTCAGAAAGATGAAGAAACA 306
QY 54 uLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
DB 307 GCAGATATCTCCCACTAGAGCTTGCAATCTGCAATCTGAAATCTTCCCTCAGAGAACCCCAATGA 366
QY 74 sasPValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
DB 367 ACATGTGCTCTCTCTTGTAGCATGTTGTTGGCTATATCTTTGCT 412

RESULT 7
US-09-918-995-33817
; Sequence 33817, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33817
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33817

Alignment Scores:
Pred. No.: 5.91e-16 Length: 433
Score: 276.00 Matches: 55
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.84% Indels: 0
DB: 11 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-918-995-33817 (1-433)
QY 1337 ThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAla 1356
DB 3 ACGGACAGAGAGTCCAAAGCAACAGCAGCAGAGTCAAGAGAGAGCAGACAGAGAGCA 62
QY 1357 GluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGly 1376
DB 63 GAATCTCTGATATGATGCAATGGAATGCAACACAGCTCCACACACAGAAAGAGAGAGAGA 122
QY 1377 ArgProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
DB 123 AGACCATCAAAAAGCCATCAACATCAACACCAAAAAAATGTC 167

RESULT 8
US-10-071-766-51
; Sequence 51, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huel-Mel Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 51
; LENGTH: 8493
; TYPE: DNA
```


Dh 2485 GGGAAACAAAGATGACCTAGCACTACACAGTCGAATTATATAAGCACTGATCAAGAA 2544
Qy 646 Ile----- 646
Db 2545 TTCCAAATTTCCAAACCCCTTCATATGACATTGAGCAAAAGCTATAGATGCTCTGAG 2604
Qy 647 -----ATGAlaGlyLeuGluLeuLeuValLeuSerPheThrHisProIleSer 663
Db 2605 GAGATGAGAGATGAATCAGGAATAGTATATCTCTGAAA----- 2646
Qy 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetasp----- 680
Db 2647 -----GAAGCCCAAAATTTGATTCGAGTTGCGTGTGAAAGCCGAGCTTCTTAC 2700
Qy 681 -----AspGluLysValAlaGluAlaLeuGluIlePheLysasnThr 695
Db 2701 AAGACCCAAAGAACTTCAGAGAAACACGTGAGTTCAAGAAAGACTA-----AATGAG 2754
Qy 696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValIleuHis 715
Db 2755 ATGGAACAAGCTGAAGGAACAATTAGAAATAGAGATTCTCCGCTGCAAACTGATAGAAAG 2814
Qy 716 HisLysSer----- 718
Db 2815 GAGAAACACGATTTACTGAGAAACTGACCAAACTTTAGAGAAGTAAAAACTTTAACT 2874
Qy 719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
Db 2875 CAGAAACAAAGATGATCTAAACAACTCCAAAGAAAGCTTGCAAAAT----- 2919
Qy 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
Db 2920 ---GAGAGGACCAAGCTCAAAAGATATGTCAGACTGTTAAACATGAATTAATTACT 2976
Qy 758 Serasn-----LeuGluHisLeuIleThrProLeuValThrIleGluHis 772
Db 2977 CAGAAACAATTACGAATATGCTGTGATCTGTGAACAAACATCAAGAAACAATTAAATACA 3036
Qy 773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTyrLysSerTyrValAlaThrPhe 792
Db 3037 CTA-----AAATCGAAATTTTCTGAGAA 3060
Qy 793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysThrThrLysLeuTyr 812
Db 3061 GTTTCACAGAAATTGTCATATG----- 3081
Qy 813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
Db 3082 -----GAGAAATAATACAGAGAAACTTAAGATGAATTTCAG----- 3117
Qy 833 ValArgTyrPheLeuGluMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
Db 3118 ---CAAAAGATGGTGGCATAGATTAATAAACAGATTTGGAAGCTAAATAATCC---CA 3171
Qy 853 LeuLeuThrThrIleLeuHisSerAspLysAspLeuThrGlnGlnGlyLysIleSerLys 872
Db 3172 ACACAACTGACAGATGTTAAGATTAATAGAGTAATTAAGCAACAAGAGATATTTCT 3231
Qy 873 -----ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
Db 3232 TTAATACAGAGAAATAATGAACTCCAAACAATGTTAGAGAGTGTATAGCAGAAAGGAA 3291
Qy 891 Gln-----GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
Db 3292 CAATTTGAAGACTGACCTTAAGGAAATATTTGAATGACCATTTGAAACACAGAAACAATTA 3351
Qy 905 GlnLeuCysAlaLeuAlaIleAsnAspLysCysTyrGlnValArgGlnValAlaPheAlaGln 924
Db 3352 AGACT-----CTTGGGAGTGAACCTTAATAAAGCAACAAGATAGTTCACACA 3399
Qy 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
Db 3400 GAAAGAACCATGCGCATTAAGAA-----GAAGGAGAGCTTTCTTAGAGCACTGTGAC 3450

Qy 945 LeuCysAla-----LysAspProValLysGluArgAlaHisAlaArgGlnCysLeu 962
Db 3451 AGACTGGCAGAGATTGAAGAAAACTTAAGAAAAAGCCCAAACTCAAGAAACAG 3510
Qy 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
Db 3511 CAAACAATCTTAATGTACACAGACAGATGATGATGATGCAAAAAAGATTATGAATTA 3570
Qy 982 -----LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
Db 3571 GAGAAATTAAGAAATGAATTAAGAAACAAAGAAATGACATTGCAACATATGGAACAGAG 3630
Qy 995 ThrIleHisLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
Db 3631 AGCTTGATTTGGCTCAGAAACTTAATGAAATTAATGAGAAATTAATCTTAATACCAAA 3690
Qy 1013 GlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsn 1032
Db 3691 GAAAGAAAGTTCTTAAGGAA-----TTACAGAACTCATTTGAAACAGAG 3735
Qy 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
Db 3736 AGAGACCACTTAGAGATATATTAAGAAATTAAGAACTAGACAGCCTCAAAACCAAGAA 3795
Qy 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
Db 3796 GAACTAATAATGCTCATATTTCACCTAAAGAAACCAAGAAACTTAATGATGAATACAGA 3855
Qy 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
Db 3856 AGAACCGTA-----TCTGAGAAAGCA----- 3876
Qy 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnAsnPheSerAsnThrLysAsnTyrLeu 1112
Db 3877 -----GCTCAATATTAATTAATCTCAGACACTTAAGAAATCCCATACCAAA-----TTA 3924
Qy 1113 ProGluIleMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
Db 3925 CAGAAAGAGATCCCA-----GTGCTTCAT 3948
Qy 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150
Db 3949 GAGGAACAAGATTAATCTGCTTAATGTGAATAAGTCAGAGACCTCAGGAACAATGAT 4008
Qy 1151 ArgMetGluThrValSerAsnAlaSerSerSerAsnProSerSerProGlyArgIle 1170
Db 4009 GAACGTGAGATTATTAACAGACAGCTCCAAACCAAGCACTCAACAACACTGGCAGAAATA 4068
Qy 1171 LysGlyArgLeuAspSerSerGluMetIleAsnHisSerGlu---AsnGluAspTyrThrMet 1189
Db 4069 -----GAAATGGAAGAGCTCAGCTGATGAATGAATAATTTCAAGCA 4107
Qy 1190 SerSerProLeuProGlyLysLysSerAspLysAspAspSerAspLeuValArgSer 1209
Db 4108 AGTCAGGAAGATTAATAATCTCTATACCAAGAAAGACAACTTAATAACGATTAAGAA 4167
Qy 1210 GluLeuGluLysProArgGlyArgLysThrProValThrGlu----- 1224
Db 4168 GCCCTTGAAAGTTAAACATGACCACTGAAAGCAATATTAAGAGAACTTTGGCTAAATTC 4227
Qy 1225 -----GlnGluGluLysLeuGlyMet-----AspAspLeu 1234
Db 4228 CAGGACTCTCAAGCAACAAGAAAGAGTCTTAATTAATGAAGAAAGAAACAAATGAACCT 4287
Qy 1235 ThrLysLeuValGlnGluGln-----LysProLys----- 1244
Db 4288 ACAAATTCGTGAGTGAATGAGCAATTCAAACCAAGATTCAGCACTACTAAGGATA 4347
Qy 1245 -----GlySerGlnArgSerArgLysArgGlyHis----- 1254
Db 4348 GAAATAGAAATGCTCGATGTGTCCAAAGACTTTCAAAGAAATGCATGATGAATGAATCT 4407

Db 1423 TTGAAGCATCAGATGATTTGGCTAAAGACTTGCACAGACCTTGTGTAGAGCCTCA 1482
 QY 361 AsprogluValAlaIleargHisasprValIleValSerIleValThrAlaAlaLys 380
 Db 1483 GACTCAAAAGGGAAC-----TCGTATGTTTATTACAGTGAAGAAACCATAGCTCAAG 1536
 QY 381 AspiIleuValAsnAsp----- 387
 Db 1537 CATAGTACACTAGTCAGCAGGTTCATGAAAAAGTCTTCTTCTTAGAAACCAAGCTTCAG 1596
 QY 388 -----HisIleuAsnPhenValArgIle----- 395
 Db 1597 GGCATTGGGCAATTCACAGAAATACCATTCGAGAAATGTTTCTCACTGCGAGAGTTGAT 1656
 QY 396 -----ArgThrLeuAspLysArgTrp 402
 Db 1657 GATGAAGTGAATAGCAGTCCAGTGGGAGAGATGCAGAAACATTGCAAAAGCAAAAG 1716
 QY 403 ArgValArgLys-----GluAlaMetMetGlyLeuAlaGlnIleThr 416
 Db 1717 GAAACTATAAAAGCCTTCTAAAGAACTAGAAAGCCTCATGTGCAAGCAATGACATGCC 1776
 QY 417 LysLysThrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIle 436
 Db 1777 AATAAAACCTGCAGATGATGTTAGCCACAGAAACCTCTCCTGACCTTGTGGATTC 1836
 QY 437 LysAsp----- 438
 Db 1837 AAAAGGCACTTGAAGCCTTAAGCAAAACATGCACAAAGTTACTGGACCGCAAGCCC 1896
 QY 439 -----LysLeuLeuHisIleThrTyrGlnAsnSerIle 449
 Db 1897 AGAGAAAGCAGGTTGAAGGACAAATTAAGCCCTTGAAGAAATTTACACCAATTTGAAA 1956
 QY 450 AspasprArgLeuValGluArgIle-----PheAlaGlnTyrMetValPro 465
 Db 1957 GAATTTCTATTCTGCTCCAGAAAGCCGAGAACATGAAGAGTCAAGAGTCCCTGTGGT 2016
 QY 466 HisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp 485
 Db 2017 ATGGAACGAGACAAATTAATACAGCACTTAACATGTTCAAGTAATTCAGAAAGAG 2076
 QY 486 Leu----- 486
 Db 2077 ATTGAACCTTGCAGAGTAACACAGCAAGATGTAACCTGTTAGTCAAGGCTTATTCAG 2136
 QY 487 AsnAlaValLysAla-----LeuAsnGlu 494
 Db 2137 AGTCTGTCAAAAGCACTAGCACTCAGGCTTGGAGCATGACCTGATGATGTCATGCA 2196
 QY 495 MetTrpLys-----CysGlnAsnLeuLeu 502
 Db 2197 CGGTGGAAGACTCTCAATAGAGGTGGCTCAGCGAGAGCCAGCTTCAGAGAGCCTTG 2256
 QY 503 ArgHis-----GlnValLysAspLeuAspLeuIle----- 513
 Db 2257 CTGACACTGTGGAGATTCCAGATGCTCCCTGAGATCCCTGCTCAGCTGATGATGGACACT 2316
 QY 514 -----LysGlnProLysThrAspAlaSer-----ValLysAlaIle 525
 Db 2317 GAGAGAGTTGTGGCCAATCAAGAACCCCGCTGGGTGAGTTCAAAAGTGTAAAGGCCAG 2376
 QY 526 PheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPhe 545
 Db 2377 ATACAAAGCAAAAGCTTCTCCAGAGATTGTTGATGACGAAAAATCTACGCTGAGAGTA 2436
 QY 546 MetLysLys-----PheThrGlnValLeuGluAspAspGluLysIle 559
 Db 2437 ATCAAAAGAGAGAGAAAAAATTCATACACACAGAGAGCCCGCAGATTAAGTGAAGTT 2496
 QY 560 ArgLysGlnLeuGluValLeuValSer-----ProThrCys 571
 Db 2497 TTGAAACAGCTCACTCTTGTGATAGCAGATGGAGGCAATGCTTAATAAGCTGAACA 2556

QY 572 SerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLys 591
 Db 2557 AGGAATCGTCAAGTTGGAAAGT---ATCTCGGTGGAGCAGACGCAATTTTCATGAACCTTA 2613
 QY 592 GlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProVal 611
 Db 2614 GAACCACTGAAAGAGTGGCTTACAAACCATAGAAAAAGAGCGTGTGAATGTGAACCCATA 2673
 QY 612 HisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGly 631
 Db 2674 GGAACCCAGCAATCTTAACCTTGAGGCAACAAATTTGCACAG---CACAAAGTTCTGCAGAG 2730
 QY 632 ThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGly 651
 Db 2731 GACATCTTAATCTCAGAAACAAATGTA-----GATCAGGCTTAATTAATGATTAGAA 2784
 QY 652 LeuLeuLys-----ValLeuSerPheThrHisProIleSerPheHis 665
 Db 2785 CTACTTAACAAACCAACAGTGCATGCAACTTTTAATTAATCAAGTAATTTGGAAGCCATT 2844
 QY 666 SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla 685
 Db 2845 AAAGCAAGGTACAAAGAC-----ATTACTAACTGACAGCACTGATGTCGCAAGACTCTG 2898
 QY 686 GluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705
 Db 2899 GACAGGCGCTGCACCTTGCAGAGCGGCTGCAGCTCAGACAGAGAGAGCTGTGAACTGG 2958
 QY 706 IleArgSerAlaLeuLeuProValLeuHisIleLysSerLysGlyProProArgGln 725
 Db 2959 CTGACCAAGTGGAGAGTGAATTAATCTTCATATGAATCTCAG----- 3000
 QY 726 AlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAla 745
 Db 3001 -----CTTGAAAGAGAGAAAGCAAGCAAGTCAACCA 3030
 QY 746 GlnIlePheGluProLeuHisLysSerLeuAspProSerGlnLeuGlnIleLeuIle-Th 765
 Db 3031 CAATGG---AGACCAAGAGAACTGMAAAGAGAAAGCAAGAAACCAAGCCTTACTGGAC 3087
 QY 765 rProLeuValThrIleGlyHisIle----- 773
 Db 3088 TCCCTTAATGAAGTGAAGAGAGCTTTCGTGGAAGCTGATGAGGAGGCAAGAAAGCA 3147
 QY 774 -----AlaLeuAlaIleProAspGlnPheAlaIlePro----- 784
 Db 3148 CTTGAAAAATGTAGTGAAGACAAATGAGCGCTAACGATTAAGTGAAGGAGACATCACT 3207
 QY 785 -----TrpLysSerTrpValAlaIleThrPhe-----IleValLysAspPhe 797
 Db 3208 CAGAAAGTGAAGAGAGAGATGCAAGCAACCTTCTGCGAGTCAAGCAAGCTTGAACCAAGCT 3267
 QY 797 uLeuMetAsnAspArgLeuProGly-----LysLysThrThrLysLeuTrpVal 813
 Db 3268 GATGCTGAGT-----TATCTGATGTTACTGAACAGAAAAAATGATGATGCTGGGT 3321
 QY 813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetVal 833
 Db 3322 GACATCAGAGCTT-GAGCAAGACCAAGACTTCTCAGAGCTTCAAGTTCAAAG----- 3372
 QY 833 lArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerLysThrIleSerThrArgGly 853
 Db 3373 -----ACATTCCACCAAGAGAT 3389
 QY 853 uLeuThrThrIleLeuHisSerAsp-----GlyAspLeuThrGlnGluGly---LysIle 870
 Db 3390 TTGAGAG-----CACAGAGATATGATGATGATGATGATGATGATGATGATGATGAT 3440
 QY 870 e-----SerLysProAspMetSerArg----- 877
 Db 3441 CATGACGCGATGACGTGAAGAGAAAGCAATCAATGAGAAAGAAAGCTGCAAGGATCT 3500

```

QY      878 -----LeuArgLeuAla1 883
Db      3501 GAAGAACTATGATACACTCTGCAGATTAAATTCAGAAAGTATGTCAGCTGGAACGGC 3560
QY      883 ySerAla1leValLysLeuAla1ngluProCysThyHISgluLe-----Ileth 900
Db      3561 ACAGTCCCTGGTTAAACCAATCTGGGAA---ACATATGAAGAACCTTTGGCCATGGCTGAC 3617
QY      900 rLeuGlu1n-----TyrGluLeuCysAla1leAla1le----- 911
Db      3618 AGAAACACATCATCATCTCTCAGCTTCCGCCGCCGCCCTTAATATGAACTCTAAG 3677
QY      912 -----AsnAspLysCysTyrGluValArg1nVal1phea1gln----- 924
Db      3678 GCAGACACAGACAGACATCGGCAACGCTGAGTTAGCTGAACACACAGCCTCATAT 3737
QY      925 -----LysLeuHISLys---GlyLeuSerArgLeuArgLeu---ProLeuGlu----- 938
Db      3738 AGATAGATGACAAACCTGGCCACAGTTACTGGAATGAGCCCTGGGGAAGGCTTTTC 3797
QY      939 -----TyrMetAla1leCysAla1leCysAla1-----LysAspProVal1 952
Db      3798 TATCCAGAGAGATGTGGCAGCCGACACCCCTTACAGTCAATTAAGAGATGTCA 3857
QY      952 sGluArgArgAlaHISAlaArgGlnCysLeuValLysAsn1LeasNValArgArg1uTy 972
Db      3858 AAAGCCGTGTGTG-----GCACGTGATGAAGCCATTTCTCATCACTCACTGATT 3905
QY      972 rLeuLysGlnHISAla1leValSerGluLysLeuLeuSerLeuLeuProGluTyValVa 992
Db      3906 CCATGACACAGATAGATAGATCCCTTGAGACCTGGAACGCCATCCTGGACGCTGAGCA 3965
QY      992 lProTyThr1LeHISLeuAlaHISAspProAspTyrValLysValGln----- 1009
Db      3966 GCCACCCCTGTATC-----TCTGCAGAGGTGAGAAAGATCAAGAACAGATCACTGA 4016
QY      1010 -----Asp1leGlu1nLeuLysAspValLysGluCysLeuTyrPh 1023
Db      4017 AAATAGAAATGTGTCACTAGACATGAGAAAGCTACAGCCGTTGTATGAACCTCTTAACA 4076
QY      1023 eValLeuGlu1leuMetAlaLys----- 1031
Db      4077 GAGGGGAGAGAAATGTGCTTACATCTGGGGGACCTGATTAAGACATATCTGCCAAGC 4136
QY      1032 -----AsnGluAsnAsnSerHISAlaPhe1leArgLysMetValGluAsn1leLysG1 1049
Db      4137 TGTTCAGATTAAGCTTACCAAAATGTTTCATTTGGGAGAACATACACACACTGGGA 4196
QY      1049 nThrLysAspAlaGln1ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCy 1069
Db      4197 AGAGAGGGAAGCCAAACTACTGTGATGTGATGACCTAGCAGAAAGTTCTGG-----TG 4250
QY      1069 sAspValAlaMetAsn1leIleMetSerLysSerThrThrTyrSerLeuGluSerProLy 1089
Db      4251 TGATCACATGTCATGTATAGTACATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 4310
QY      1089 sAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnHeserAsnThrLy 1109
Db      4311 AGATCCCGAATGATTCCTTCAGTAAACAAACACACAGAACACACAGACACATTAAG 4370
QY      1109 s-----AsnTyrLeuProProGluMetLysSerPhePheThr----- 1121
Db      4371 GGAAGAAATAGATGACTACAGAGAGCTGATATGTTAATTAACCTAGATTCTGAAC 4430
QY      1122 -----ProGlyLysProLysThrThrAsnValLeuGlyAlaValas 1135
Db      4431 CATTCGGCGCATGTGGGAGCTGATTAACCATTTGTCAAGAAAGATAGTAGTTAA 4490
QY      1135 nLysProLeuSerSerLacGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVa 1155
Db      4491 TTCAGCATGGGATTCCTATAATAAGCTTGAAGAACCGGATTCACAAACTTGGAGAGCC 4550
QY      1155 lSerAsnAlaSer----- 1159

```

```

Db      4551 AATGCAAGCCGCTTCAGTACCAGATGACATGACGGCGGTATTGACTGGCTAGATAT 4610
QY      1160 -----SerSerAsnProSerSerProGlyArg1leLysGlyArg1e 1174
Db      4611 TGCAGGTGTAATTACCTTCAATGTCTCCAAATTTGAACAGATCTGAAACTGTCAAGCA 4670
QY      1174 uAspSerSerGluMetAspHISserGluAsnGluAspTyrThrMet----- 1189
Db      4671 GCAGATTGAAGAGCTTAAGCAATTTAAGTCTGAGGCCCTATCAACAGCAGATAGAAATGCA 4730
QY      1190 -----SerSerProLeuProGlyLysLysSerAspLysArgAspSerAs 1205
Db      4731 AAGACTGAATATCAACAGCAGAGAGCTTTGCTAAAGAAAGTAAACAGAGAGAGTCAACAA 4790
QY      1205 pLeuValArgSer-----GluLeuGluLysProArgGlyArgLysLysThrProVa 1222
Db      4791 CACTGTTCAAGACCCATTAATGTGAACTGAAATTTGATATGGGATAGCCTGAGAGAGAAAT 4850
QY      1222 lThrGluGlnGluLysLeu----- 1230
Db      4851 CATCAACAGACAGCATTAACCTGGAGGGTGTCTATTAGCCTGGGTCAGTTCCAACTGC 4910
QY      1230 yMetAspAspLeuThrLys-----LeuValGlnGluGlnLysPr 1243
Db      4911 CCTGATGAGACTCTGTGCATGGCTGACACACACCGAGGCTTGTCTAAGTACAGAAACC 4970
QY      1243 oLysGlySer----- 1246
Db      4971 TGTTTGGAGAGACCCCTTAACCATTTGAATTTGAACCTTGCACACATCATGCTCCAAA 5030
QY      1247 -----GlnArgSerArgLysArgGlyHISThrAlaSe 1257
Db      5031 TGATGTATTAAGCCATCATGTCACAGTGAAGCCGTTAATAAGACAGAAATGATCTAAT 5090
QY      1257 rGluSerAsp-----GluG1 1262
Db      5091 TGAATCAAGTGCACAGAGAGAGACAGCAACCTTCACAGACAGCTAGAGTTTAAATCA 5150
QY      1262 nGlnTyrProGlu1uLysArgLeuLysGluAsp1leLeuGluAsnGluAspGlu1nAs 1282
Db      5151 ACAGCTG-----CAAAATGTTTGGAAAAAACAACAAAGGAA 5189
QY      1282 nSerPro-----ProLysLysGlyLysArgGlyArg----- 1292
Db      5190 GCACACCTGATGATGCTCTTGCGCCAGCCAAAGGGTTCATTAACCCGCTGGAGGATT 5249
QY      1293 -----ProProLysProLeuGlyGly-- 1299
Db      5250 GCAGCACTGGCTGACTGACACAGAGCGCTCATCTGTTGGCATTAACCCGCTGGAGGATT 5309
QY      1300 -----GlyThrProLysGlu----- 1304
Db      5310 ACCGGAACACCCAGAGAGCAAGCTTAATGTCCATATGAAGTCTGTCTCCTTTGAAGC 5369
QY      1305 -----GluProThrMetLysThr---SerLysLysGlySerLysLysSerGlyProp 1322
Db      5370 TAAAGAGAAATATTAAGAGTGTGATGACGAAGAGCCACACAGATGCTTCAGAGATCCC 5429
QY      1322 oAlaProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1342
Db      5430 A---AAATCTGCAGAGCAATATATTGACCAAGACATTAATTAACCTTGAAGAAAAATGGA 5486
QY      1342 sSerLysGlnHISArgValSerArgArgAlaGlnArgAlaGluSer 1358
Db      5487 ATCGGTGGAACCAAACTCAATGAAGAAACTAACTGAAGAGGCT 5535

```

RESULT 10
 US-10-037-270-130
 ; Sequence 130, Application US/10037270
 ; Publication No. US20030104529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom

/ APPLICANT: Liu, Chenghua
 / APPLICANT: Asundi, Vinod
 / APPLICANT: Zhang, Jie
 / APPLICANT: Ren, Feiyan
 / APPLICANT: Chen, Rui-hong
 / APPLICANT: Zhao, Qing A.
 / APPLICANT: Wehrman, Tom
 / APPLICANT: Xue, Aidong J.
 / APPLICANT: Yang, Yonghong
 / APPLICANT: Wang, Jian-Rui
 / APPLICANT: Zhou, Ping
 / APPLICANT: Ma, Yuning
 / APPLICANT: Wang, Dunrui
 / APPLICANT: Wang, Zhiwei
 / APPLICANT: Tillinghast, John
 / APPLICANT: Dimanac, Radoje T.
 / TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
 / FILE REFERENCE: 784CIP2B
 / CURRENT APPLICATION NUMBER: US/10/037,270
 / PRIOR FILING DATE: 2002-01-04
 / PRIOR APPLICATION NUMBER: 09/552,317
 / PRIOR FILING DATE: 2000-04-25
 / PRIOR APPLICATION NUMBER: 09/488,725
 / PRIOR FILING DATE: 2000-01-21
 / NUMBER OF SEQ ID NOS: 1104
 / SOFTWARE: PL-FL-genes Version 1.0
 / SEQ ID NO 130
 / LENGTH: 8503
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (91)..(8082)
 / US-10-037-270-130

Alignment Scores:
 Pred. No.: 7,36e-10 Length: 8503
 Score: 231.00 Matches: 285
 Percent Similarity: 35.41% Conservative: 267
 Best Local Similarity: 18.28% Mismatches: 541
 Query Match: 3,21% Indels: 467
 Gaps: 67

US-09-512-581b-2 (1-1391) x US-10-037-270-130 (1-8503)

QY 17 ProGluValLysGluIleSerAspLysIleSerIysGluGluMetValArgArgLeuLys 36
 DB 1087 CCTATGTTAATGAGGTATCA-----ACTGATGAAGCTCTCTGAAAAAGTATAGA 1137
 QY 37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
 DB 1138 -----AAAGAAATATGATGATCTTAAAAACAATTAGAGAGAGGTTTCTTTAGAGACG 1188
 QY 52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
 DB 1189 CGGGCTCAGCAATGGAAGAAAGCAATTTGGCCCACTTTTGGAAAGAA-----AAAGAT 1242
 QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
 DB 1243 TTGCTTCAGAAATGACAGAAATTTGAAAACTTAACACGCGATCTGGTG----- 1296
 QY 83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
 DB 1297 -----ACCTCTTCTTCC----- 1308
 QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
 DB 1309 -----CTCAGCTTGCAG-----CAGGAATTAAGAGCT 1335
 QY 123 ProGlnPheAsnArgTyrPheTyrLeuLeuGlnAsnIleAlaThrValLysSerTyrAsn 142
 DB 1336 AAAAGAAAAAGCAAGTACTTGTGCTTGGCAAAATTAACAAATGAAAGAACTCAAC 1395

QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrIleup 162
 DB 1396 -----TATGCAAGATCAATTT 1410
 QY 163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
 DB 1411 AATATACCAACAAT---ATACACACAAAAACACATTAAGCTTTCTATGAATATTAATACGA 1467
 QY 183 SerIle-----IleCysGluGlyAspThrValSerGlnGluLeuAspThrVal 199
 DB 1468 GAATGTGATGATCTGTCTGTTCAGAGCTGTGATGTTTTCAGTAACACCTTGATCATCTGA 1527
 QY 200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
 DB 1528 AGTGAGATGAAGATGCAACCAACCAAGCTACTAAATGACGAGAAATATAGAA----- 1581
 QY 219 LysAlaLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238
 DB 1581 ----- 1581
 QY 239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeu 258
 DB 1582 -----AGTACGTTGAACCTCCTTCTGCTGACTATGATATCTCG 1620
 QY 259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuLys 278
 DB 1621 GTATTA-----GACATATGACAACTTCACACAGCAAGAAAAAAATATGGAA 1665
 QY 279 PheLysLeuLysSerAsnAspAsnGlnGluArgLeuGlnValValLysLeuLeuAlaLys 298
 DB 1666 TTGAATTTTAAAGAAAGAAATGATTTGGATGATGATTTTGAAGCGCTTGAAGAAAGAACTTAA 1725
 QY 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyr 318
 DB 1726 -----AAAGATCAAGACATGCCAA----- 1743
 QY 319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
 DB 1743 ----- 1743
 QY 339 HisCysLeuMetAsnHisProAspPheValLysAspLeuThrGluTyrLeuLysValArg 358
 DB 1744 -----CTAATTCATGAAATTTTTCGAACCTTAAAGATTTAGTAAACATCGAAGTATAT 1797
 QY 359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
 DB 1798 AATCAAGATCTTGGAGAAATGAACTC-----AGTTCA 1827
 QY 379 LysLysAspIleLeuLeuValLysAspHisLeuAsnPheValArgGluArgThrLeu 398
 DB 1828 AAAGTAGAGCTGTAGAGAAAGAACACAGATTAAGAAAGCTTACAGAGAAATACATACAC 1887
 QY 399 AspLysArgTyrArgValArgLysGluLysIleMetMetGlyLeuAlaGlnIleTyrLysLys 418
 DB 1888 TCTCAAAAG-----CTAGAAATATTAATAATGAGCTGTGCA----- 1923
 QY 419 TyrAlaLeuGlnSerAlaIleGlyLysAspAlaLysGlnIleAlaThrIleLysAsp 438
 DB 1924 TACTCATTTGGAAGCATTTGAAGACCCAAACAAATGAAGACGACTCTTGTGATGCTGAA 1983
 QY 439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
 DB 1984 ACTGTA-----GCCCTTGATGCCAAGCA 2007
 QY 459 PheAlaGlnTyrMetValProHisAsnLeuGluThrGlnGluArgMetLysCysLeuTyr 478
 DB 2008 GAATCAGGCTTTCTTAGAAGTGAATGTGATGGAAGAGAAATGAAGAACTTGCA 2067
 QY 479 TyrLeuTyrAlaThrIleAspPheLysAlaValLysAlaLeuAsnGluMetThrLysCys 498
 DB 2068 ACTACATACACAGCAAAATGGA-----AATGATATTCAGCTTATAT 2106

QY 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeuLysGlnProLysThr 518
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:|
 Db 2107 CAAGCCCAATGGAGGCAAAAGAAATGCAGATTGATCTCGAAGAAAGAA----- 2157
 QY 519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg----- 534
 2158 -----TTACATATCTGCTTTTAATGAGATTAACAACAACTCACCTCCCTTANAGANGGC 2208
 QY 535 -----AsnLeuProAspProGlyLysIleGlnAspPheMet 546
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:|
 Db 2209 AAGTTCCAAAGATTTGCTGCTGTAATTTGGAAATGGAGAAAGATTACTGATCTTCAG 2268
 QY 547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
 |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2269 AAGAACTAATTAAGAAGTTGAAGAAATGAAACCTTTCGCGGAAGAAGTC---ATTTCG 2325
 QY 567 ValSerProThrCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLys 586
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2326 CTTTCA-----GAATTGAATCTTTACCTTCTGAAAGTAGAAAGCGTGAAGAAAGAG 2376
 QY 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2377 ATA-----CAAGCAAAATCTGAGAGCTCCATTAATTAACATCAGAAAGAAAGAT 2424
 QY 607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
 |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2425 AAATTTGTTTCTGAAAGTAGTTCATTAAGAGAGTAGAGTTCAAGGTTTACTTGAAGAAT 2484
 QY 626 AsnLysSerIleAspGluThrAlaAspAspGluAspLysValProThrAspAla 645
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2485 GGGAAACAAAGATGACTAGCAACTACACAGTCGAATTATTAAGACACTGATCAAGAA 2544
 QY 646 Ile----- 646
 Db 2545 TTCCAAATTTCAAACCTTCATATGAGCTTTGAGCAAAAGTATAGATGCTCTGAG 2604
 QY 647 -----ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProLysSer 663
 |||:|:| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2605 GAGAAATGAGAGATGAATCAGAAATAGTTAATCTCTCTAA----- 2646
 QY 664 PheHisSerIleGluThrPheGluSerLeuLeuLysLeuLysMetAsp----- 680
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2647 -----GAAGCCCAAAATTTGATTCGATTGGGCTTTGAAAGCCGAGCTTCTTAC 2700
 QY 681 -----AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThr 695
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2701 AAGCCCAAGACTCTCGAGAAACACGTCGAGGTTCAAGAAAGACTA-----AATGAG 2754
 QY 696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2755 ATGAGACAGCTGAAGGAACAATTAGAAATAGAGATTCTCCGCTGCAAACTGTGAAGAG 2814
 QY 716 HisLysSer----- 718
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2815 GAGAAACACTGATTACTGAGAAACCTGCAGCAAACTTTAGACAGAAAGTAAAACTTTACT 2874
 QY 719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2875 CAGAAAGAAATGATGTAAACAACTCCAGAAAGCTTGCAAAAT----- 2919
 QY 738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2920 ---GAGAGGAGCAAACTCAAAAGATGATATTCAGGATCTGTAAACATGAAATGATGACT 2976
 QY 758 SerAsn-----LeuGluHisLeuIleThrProLeuValIleThrIleGlyHis 772
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 2977 CAAGACAAATTAGCAAAATGCTTGAAGTCTGTAACACATCAGAAACAAATTAATACA 3036
 QY 773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPhe 792
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3037 CTA-----AAATGCAAAATTTCTGAGAGAA 3060
 QY 793 IleValLysAspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuTrp 812

Db 3061 GTTCCCGGAATTTGCATATG----- 3081
 QY 813 ValProAspGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3082 -----GAGGAAATACAGAGAAACCTAAAGATCAATTTGAG----- 3117
 QY 833 ValArgThrLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3118 ---CAAAAGATGGTTGGCATATGATTAATAAACAGAGATTGGACGTAAAAATACC---CA 3171
 QY 853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3172 ACACTAATCTCAGATGTAAAGATTAAGATATATTCAGCAACAAAGGAATATTTCT 3231
 QY 873 ---ProAspMetSerArgLeuArgLeuAlaIleLysSerAlaIleValLysLeuAla 890
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3232 TTAAATACAGAGAAATTAAGACTCCAAACAAATGTTAGAGAGTGTATAGCAGAAAGAA 3291
 QY 891 Gln-----GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3292 CAATTTGAAGACTGACTAAAGAAATATTTGAATGACATTTGAAACCAAGAAATTA 3351
 QY 905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3352 AGACTT-----CTTGGGAGTAGAACTTAAGAACACAGAGATGTTGCACAA 3399
 QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluLysThrAlaIleCysAla 944
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3400 GAAAGAACCATGCCATTAAGAA-----GAGAGACAGCTTTCTAGACCTGTGAC 3450
 QY 945 LeuCysAla-----LysAspProValLysGluArgArgAlaHisAlaArgLysLeu 962
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3451 AGACTGCGACAGATGTGAAGAAATTAAGAAAGAACCCAGCACTCAAGAAAGAAACAG 3510
 QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3511 CAACACTCTTTATGTACAAAGAAAGATGAGTGCAGTGCAGAAAGAAATTAATGAATA 3570
 QY 982 -----LysLeuLeuSerLeuLeuProGluTyrValIleValProTyr 994
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3571 GAGCAATTTAAAGATTAATTAAGAACAAAGAAATTTGACATTTGAGCACTTAAGAAACAG 3630
 QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3631 AGGCTTGATGGTGCAGAAACTTAATGAAGAAATTAATGAGAAAGTCAATTAACCAAA 3690
 QY 1013 GlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLysAsn 1032
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3691 GAAAGAAAGCTTCTAAAGGAA-----TTACAGAAAGTCAATTTGAAACAGAG 3735
 QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3736 AGAGACCACTTAGAGATATATTAAGAGAAATTTGAACCTACAGCCCTACAAACCAAGAA 3795
 QY 1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3796 GAACATAAAATTCGTCATATTCACCTTAAGAAACCAACCAAGAACTATGATGAACATAGA 3855
 QY 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3856 AGAAGCCTA---TCTGAGAGAGCA----- 3876
 QY 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnProPheSerAsnThrLysAsnTyrLeu 1112
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3877 -----GCTCAATAATTAATTAATCTCAGAGCTTTGAAAAATCCCATACCAAA-----TTA 3924
 QY 1113 ProProGluMetLysSerPhePheThrProGluLysProLysThrThrAsnValLeuGly 1132
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
 Db 3925 CAAGAGAGATCCCA-----GTGCTTCAT 3948
 QY 1133 AlaValAlaLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150
 |||:|:| ||| ||| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

```

Db      3949 GAGGACAAAGATTACGCTTAATGTGAAAAAAGTCAGTACGAGCTCAGAGAAACATGCAAT 4008
Qy      1151 ArgMetGluThrValSerAsnAlaSerSerSerSerSerProGlyArgIle 1170
Db      4009 GAACGGAGGATTATTAAACAGAACAGTCACCAACCAAGCACTCAACCACTGGCAAGATA 4068
Qy      1171 LysGlyArgLeuAspSerSerSerGluMetAspHisSerGlu---AspGluAspTyrThrMet 1189
Db      4069 -----GAAATGGAAAGGCTCAGCGTTGAATGAAAAATTTGCAAGA 4107
Qy      1190 SerSerProLeuProGlyLysLysSerAspLysArgAspSerSerAspLeuValArgSer 1209
Db      4108 AGTCAGAGAGATTAATAATCTCTTAACCAAGAAAGCAACCTTAATAACCGATTAAAGAA 4167
Qy      1210 GluLeuGluLysProAlaGlyArgLysLysThrProValThrGlu----- 1224
Db      4168 GCCCTTGAAAGTTAAACATGACGAGCTGAAAGAACATATTAGAGAAACTTGGCTAAATC 4227
Qy      1225 -----GluGluGluLysLeuGlyMet-----AspAspLeu 1234
Db      4228 CAGAGCTCTCAAGCAACCAAGAACAGCTCTTAATATGAAAGAAAGAAAGACATGAACT 4287
Qy      1235 ThrLysLeuValGluGluGln-----LysProLys----- 1244
Db      4288 ACCAAATCTGAGTGAATGAGATGAGCAATTCAAACCCAAAGATTTCAGCACTACTAGATA 4347
Qy      1245 -----GlySerGlnArgSerArgLysArgGlyHis----- 1254
Db      4348 GAAATAGAAATGCTCGATGTGTCMAAAGACTTCMAAAGAACATGATGAAATGAAATCT 4407
Qy      1255 ThrAlaSerGluSerAspGluGluGln-----TrpProGluLys 1268
Db      4408 GTACCTAAGAGGAAAGATGACCTACAGAGGCTGCAGAACTTCTTCAATCTGAAAGTGAC 4467
Qy      1269 ArgLeuLysGluAspIle-----LeuGluAsnGluAsp----- 1279
Db      4468 CAGCTCAAGAAAGAACATTAAGAAATTTAGCTTAACACCTGGAACAGTGAAGAGAACT 4527
Qy      1280 -----GluGluAsnSerProProLysLysGlyLysArgGly 1291
Db      4528 AAAGTGTCTCATTTGCTGCTGAAAGAACAGAACTATTATGAGTTAAGACTGAAT 4587
Qy      1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
Db      4588 CTTTCAGAGAAAG-----GAACTGAAATATATCAACCATTCATAAAGCAG 4629
Qy      1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluGluGluGluGlu 1331
Db      4630 TTAGAGCAATCATGATTAATTTACAGAAACAGATCCAAAGATTTATGAGAAAGGAA 4689
Qy      1332 -----ArgGlnSerGlyAsnThrGluGlnLys---SerLysSerLysGlnHis 1346
Db      4690 CAACCTTAATATTAACAAATTAAGTGAAGGTTTCAGGAAAAAGTCAAGCTGAACAAATTC 4749
Qy      1347 ArgValSerArgArgAlaGlnAlaGlnAla---GluSerProGluSerSerAlaIleGlu 1365
Db      4750 AAGAGCACTCGCAAAAGCCAAAGATTCACGACTACAAATATAGAAAGTAAAGTCTGAG 4809
Qy      1366 SerThrGlnSerThrProGlnLysGlyArg----- 1375
Db      4810 TT-GACCAACAGACTTCAGAAAGTCAAGAAATTAACAATTATGATTAAAGAAAGAA 4868
Qy      1376 -----GlyArgProSerLysThrProSerProSerGlnProLys 1388
Db      4869 GGAATGAAAGAGTACAGAGAGGCCCTTCAGATGACAGAGAGCCAACTGAAAGAA 4925

```

```

RESULT 11
US-09-893-519A-140
; Sequence 140, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig

```

```

; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESTIVA, Thomas
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/16548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 7992
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/215005.1
; DATABASE ENTRY DATE: 1993-01-10
; RELEVANT RESIDUES: (1)..(7992)
US-09-893-519A-140

Alignment Scores:
Pred. No.: 1,02e-09 Length: 7992
Score: 229.00 Matches: 285
Percent Similarity: 35.61% Conservative: 268
Best Local Similarity: 18.35% Mismatches: 546
Query Match: 3.18% Indels: 455
DB: 11 Gaps: 66

US-09-512-581B-2 (1-1391) x US-09-893-519A-140 (1-7992)
Qy      17 ProGlyValLysGluIleSerAspLysLysSerLysGluMetValArgArgLeuLys 36
Db      997 CCTTATGTTAATGAGGTATCA-----ACTGATGAAGCTCTCCGAAAGGATATGA 1047
Qy      37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
Db      1048 -----AAGAAATATGATGATCTTAATAAACAATTAAGAGAGGTTTCTTTAGAGACG 1098
Qy      52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db      1099 CGGGCTCAGGCATATGAAAGAAAGCCAAATTTGGCCCAACTTTTGGAAAGAA-----AAAGAT 1152
Qy      67 PhePheLeuLysShsProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
Db      1153 TTGCTTCAGAAAGTACAGAAATGAGAAATTAAGAACTTAACACGAGATCTGCTG----- 1206
Qy      83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
Db      1207 -----ACCTCTTCTTCC----- 1218
Qy      103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
Db      1219 -----CTCAGCTTGCA-----CAGGATTAAGGCT 1245
Qy      123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTyrValLysSerTyrAsn 142
Db      1246 AAAAGAAACGACAGCTTACTTGTGCTTGGCAAAATTAACAATAAGAAAGTCAAAAC 1305
Qy      143 IleCysPheGlnLeuGluAspSerAsnIlePheThrGlnLeuTyrArgThrLeuPhe 162

```

D	b	1306	-----			-TATCGAGATCAATTT	1320
O	y	163	serValIleasnAsnGlnHisAsnGlnIlyValHisMetHisMetValAspLeuMetSer				182
D	b	1321	AATATACCACAAAT	-----	ATAACACAAACAAACATTAAGCTTCTTAATTTATATACCA	1377	
O	y	183	SerIle-----		IleCysGluGlyAspThrValSerGlnIleuLeuAspThrVal	199	
D	b	1378	GAATATGATCAATCTGCTGTCTGAGACTGATGATTTTTCAGTAACACTCTTGATACATTA	1437			
O	y	200	Leu----	ValAsnLeuValProIleHisLysAsnLeuAsnLysGlnIleValTyrAspLeuAla	218		
D	b	1438	AGTAGAGATAGAGATGGAATCCAGCAACAAACTACTAATCAGAGATATAGAA-----	1491			
O	y	219	LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPheAsn	238			
D	b	1491	-----				1491
O	y	239	GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeu	258			
D	b	1492	-----	AGTAGCTGAGCTCACTCTGCTGCTGACTATGATATATCTG	1530		
O	y	259	IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu	278			
D	b	1531	GTATTA-----	GACTATGAACAACACTCGACAGAAAAAGAGAAATGAGAA	1575		
O	y	279	PheLysLeuLysSerAsnAspAsnGluArgLeuGlnValValLysLeuLeuAlaLys	298			
D	b	1576	TTGAAATTTAAAGAAAGAAATGATTTGGATGAATTTGAGCCTTAGAAAGAAACTTAA	1635			
O	y	299	MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrGlnCysTyr	318			
D	b	1636	-----	AAAGATCAAGAGATGCAA-----	1653		
O	y	319	LeuGlyArgPheAsnAspIleHisValProIleArgLeuGlnCysValLysPheAlaSer	338			
D	b	1653	-----				1653
O	y	339	HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrIleLysValArg	358			
D	b	1654	-----	CTAATTCATGAAATTTTGCAGCTTAAAGCAATTTAGTTAGCAATGCAGCAAGATATAT	1707		
O	y	359	SerHisAspProGlnGluAlaIleArgHisAspValIleValSerIleValThrAlaAla	378			
D	b	1708	AATCAGATCTTGGATGAACTC-----	AGTTCA	1737		
O	y	379	LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu	398			
D	b	1738	AAAGTAGAGCTGTAGAGAAAGAGAACCATTAAGAACCTACAGAAATTCATTAAGAC	1797			
O	y	399	AspLysArgThrPargValArgLysGluAlaMetMetCysLeuLeuGlnIleTyrLysLys	418			
D	b	1798	TCTCAAAAG-----	CTAGAAATATATAATGACTTGTC	1833		
O	y	419	TyrAlaLeuGlnSerAlaAlaIleGlyLysAspAlaAlaLysGlnIleAlaThrIleLysAsp	438			
D	b	1834	TACTCATTTGGAAACATTTGAAGACCCAAACAAATAGACAGACCTGTTTGATGCTGGA	1893			
O	y	439	LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle	458			
D	b	1894	ACTGTA-----	GCCCTTGATGCCAAGACA	1917		
O	y	459	PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGlnArgMetLysCysLeuTyr	478			
D	b	1918	GAATCGACGCTTTCTTACAGTAAATCTGAGCTTGAAGGAAATGAAAGAACTTGA	1977			
O	y	479	TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnLysMetThrLysCys	498			
D	b	1978	ACTACATTAACAAGCAATGCAA-----	AATCATATTCAGTTATAT	2016		
O	y	499	GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr	518			
D	b	2017	CAAAAGCCAAITGGAGCGCAAAAAAGCAAAATGCAATGATCTGAGAAAGAA-----	2067			
O	y	519	AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg-----	534			
D	b	2068	-----	TTACAAATCTGCTTTTAAATGAGATTAACAAACACTCACCTCCCTTATAGATGC	2118		
O	y	535	-----	AsnLeuProAspProGlyLysAlaGlnAspPheMet	546		
D	b	2119	AAAGTTCCAAAGATTTGCTGTATATTTGGAAATTTGGAAGAAAGATTAATCTGATCTTCAG	2178			
O	y	547	LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu	566			
D	b	2179	AAAGAACTAATATAAGTAAGTTGAAGAAATGAACTTTTCGGGAAGAAATCTC-----	2235			
O	y	567	ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys	586			
D	b	2236	CTTTCA-----	GATTTGAATCTTTTACCTTCCTGAAAGTGAAGAAAGCTGAGAAAGAG	2286		
O	y	587	LeuGlnAsnProLysGlnProThrAsnProPheLeuGlnMetIleLysPheLeuLeuGlu	606			
D	b	2287	ATA-----	CAAGCAAAATCTGAAAGACCTCCAAATATATAATCAATCAAGAAAAAGAT	2334		
O	y	607	ArgIleAlaProValHisIleAspThrGlnSer-----	IleSerAlaLeuIleLysGlnVal	625		
D	b	2335	AAATGTTTTCTGAAAGTAGTTCAATTAAGAGAGATGAGTTCAAGGTTTACTTTGAAGAAAT	2394			
O	y	626	AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla	645			
D	b	2395	GGGAAACAAAGATGACCTTAGCACTACACAGTCAATTTATTAAGCACTGATCAAGAA	2455			
O	y	646	Ile-----				646
D	b	2455	TTCCAAATTTCAAAACCTTCATATGAGACTTTGAGCAAAAGATTAAGATGCTGCTTGA	2514			
O	y	647	-----	ArgAlaGlyLeuGlnLeuLeuLysValLeuSerPheThrHisProIleSer	663		
D	b	2515	GAGAAATGAGAAATGAATCAGAAATTAATCTGCTGTA	2556			
O	y	664	PheHisSerAlaGluThrPheGlnSerLeuLeuAlaCysLeuLysMetAsp-----	680			
D	b	2557	-----	GAAAGCCCAAAATTTGATTTGAGATTTGGCTTGAAGACGACGCTTCTTAC	2610		
O	y	681	-----	AspGluLysValAlaGlnAlaAlaLeuGlnIlePheLysAsnThr	695		
D	b	2611	AAGACCCCAAGACTTCAGAGAAACACGAGAGCTTCAGAAAGACTA-----	AATGAG	2664		
O	y	696	GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis	715			
D	b	2665	ATGGAACAGCTGGAAGCAATTAAGAAATAGATTCGCGTGCAGAACTGTAGAAAG	2724			
O	y	716	HisLysSer-----				718
D	b	2725	GAGAAACACTGATTTACTGGAACATGCGACGAAACTTTGAAAGAACTTAATCTTA	2784			
O	y	719	-----	LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe	737		
D	b	2785	CAAGAAATAAGATGATCTTAATAACAATCCCAAGAAAGCTTGCAAT	2822			
O	y	738	SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro	757			
D	b	2830	-----	GAGAGGAGCAACTCAAAATGATGATTCACAGATACATGATGATATGATACT	2886		
O	y	758	SerAsn-----	LeuGlnHisLeuIleThrProLeuValThrIleGlnHis	772		
D	b	2887	CAGACAAATTAAGAAATGCTCTTGAGCTCTGTAACAACATCAAGAAACAAATTAATCA	2946			
O	y	773	IleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPhe	792			
D	b	2947	CTA-----	AAATCGAAATTTCTGAGGA	2970		
O	y	793	IleValLysAspLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThr	812			
D	b	2971	GTTTCCAGGAATTTGCTATG-----	2991			

QY	813	ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMet	832
Db	2992	-----GAGGAAATATACAGAGAAATCAATAGATGAATTTCCAG-----	3027
QY	833	ValArgThrLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg	852
Db	3028	---CAAAAGATGGTTGGCATATGATAAAAAACAGATTTTGAAGCTAAAAAATAC---CAA	3081
QY	853	LeuLeuThrThrIleLeuHisSerAspLysAspLeuThrGlnGlnGlyLysIleSerLys	872
Db	3082	ACACTAACTGCAGATGTTAAAGCATTAATACGATTAATTCAGCAACAAGAAACATATTTCT	3141
QY	873	-----ProAspMetSerArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAla	890
Db	3142	TTAATACAGAGAAAAATGAAGCTCCAAACAATGTTAGAGATGGTATTCAGAAAAAGSAA	3201
QY	891	Gln-----GluProCysIleThrIleGlnIleIleThrLeuGlnGlnIleThr	904
Db	3202	CAATTTGAAGAGCTGACCTAAAGGAAATATTTGAAATGACCATTTGAAACACAGAAATTA	3261
QY	905	GlnLeuCysAlaIleLeuAlaIleAsnAspGluCysIleCysIleValArgGlnValPheIleGln	924
Db	3262	AGACT-----CTTGGGGATGACCTTAAAAAGCAACAAGATATGTCACAA	3309
QY	925	LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnIleCysAlaIleCysAla	944
Db	3310	GAAGAAAGAACATGCCATTAAGAAA-----GAGGAGAGAGCTTTAGAGACCTGTGAC	3366
QY	945	LeuCysAla-----LysAspProValLysGlnArgAlaIleHisAlaArgGlnCysLeu	962
Db	3361	AGACTGGCAGAGTTGAAAGAAAACCTAAAGGAAAAGACCCAGCAACTCCAGAAAACACAG	3420
QY	963	ValLysAsnIleAsnValArgArgGluIleThrLeuLysGlnHisAlaIleValSerGlu---	981
Db	3421	CAACAACCTCTTAATGTACACAAAGAGATGAGTGAGATGCCAGAAAACATTAATGAATA	3480
QY	982	-----LysLeuLeuSerLeuLeuProGluIleValValProThr	994
Db	3481	GAGATTTAAAGATGAATTAAGAACAAAGAAATTTGACATTTGAAACATATGGAACACAG	3540
QY	995	ThrIleHisLeuLeuAlaHis---AspProAspPheValLysValGlnAspIle---Glu	1012
Db	3541	AGCGCTTGAAGTTGGCTCAGAAACCTTAATGAAATTAATATGAGGAAGAAATCATATTAAC	3600
QY	1013	GlnLeuLysAspValLysGlnCysLeuThrPheValLeuGlnIleLeuMetAlaLysAsn	1032
Db	3601	GAAGAAAGAAAGTTCTTAAGGAA-----TTACAGAGATCATTTTGAACACAG	3645
QY	1033	GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp	1052
Db	3646	AGAGACCAACCTTAAGAGATATATATAGAAATTTGAAGTACAGGCTTACAAACCAAAAGAA	3705
QY	1053	AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuIleThrValCysAspValAla	1072
Db	3706	GAACATAAAATATGCTCATATTCACACTAAAGAACACCAAGAAACATATGATGAACATA	3765
QY	1073	MetAsnIleIleMetSerLysSerThrThrIleThrSerLeuGlnSerProLysAspProVal	1092
Db	3766	AGAGGCGCA-----TTCGAGAACAGCA-----	3786
QY	1093	LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnThrLeu	1112
Db	3787	-----GCTCAATATATTAATATACAGAGACTTAAAGAAATATCCCATACCAAA-----	3834
QY	1113	ProGlnGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGln	1132
Db	3835	CAAAAGAGATCCCA-----GTGCTTCAT	3858
QY	1133	AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser	1150
Db	3859	GAGCAACAAGATTACTGCTTAATATGAAAAAATCATGATGAGCTTCAGGAACAAATGAT	3918

[illegible]

```
; FILE REFERENCE: PA-0049 US
; CURRENT APPLICATION NUMBER: US/10/133,013
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/287,067
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ. ID NOS: 271
; SOFTWARE: PERL Program
; SEQ. ID NO 260
; LENGTH: 8491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20030166903A1 441283.5
US-10-133-013-260

Alignment Scores:
Pred. No.: 1.12e-09 Length: 8491
Score: 229.00 Matches: 285
Percent Similarity: 35.61% Conservative: 268
Best Local Similarity: 18.35% Mismatches: 546
Query Match: 3.18% Indels: 455
DB: 12 Gaps: 66

US-09-512-581b-2 (1-1391) x US-10-133-013-260 (1-8491)
QY 17 ProGlyValLysGluIleSerAspLysIleSerLysGluLumetValArgArgLys 36
DB 1087 CCTATGTATATGAGATATCA-----ACTATATAGCTCTCCGAAAAGGTATAGA 1137
QY 37 MetValAllysthrPheMetAspMetAspGlnAspSerGluLys----- 51
DB 1138 -----AAAGAAATATATGATCTTAAAAACAATTAAGACGAGCTTCTTAGACAGC 1188
QY 52 -----GluLysGluLeuTyrlLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
DB 1189 CGGGCTCAGGCATGGAAGAAAGACCAATTGGCCCACTTTTGGAGAA-----AAAGAT 1242
QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
DB 1243 TTGCTTCAGAAAGATACGATGAGAAATTTGAAAACCTTAACAGCATGCGGTG----- 1296
QY 83 CysLeuAlaAspIlePheArgIleTyrlAlaProGluAlaProTyrlThrSerProAspLys 102
DB 1297 -----ACCTCTTCTTCC----- 1308
QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuGluAspThrLysSer 122
DB 1309 -----CTCACGTTGCAA-----CAGGAATTTAAAGGCT 1335
QY 123 ProGlnPheAsnArgTyrlPheTyrlLeuLeuGluAsnIleAlaTrpValLysSerTyrlAsn 142
DB 1336 AAAAGAAAGACGAGATTACTTGCTGGCCTTGGCAAAATTTAAACAATAAGAACTCAAC 1395
QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrlArgThrLeuPhe 162
DB 1396 -----TATGCAAGATCAATTT 1410
QY 163 SerValIleAsnAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
DB 1411 AATATACCAACAACAAAT-----ATAACACAACAAACATTAAGCTTTCTATAAATTTATTACGA 1467
QY 183 SerIle-----IleCysGluGluLysPThrValSerGlnGluLeuLeuAspThrVal 199
DB 1468 GAATATGATGAAATCTGCTGTTCAGAGTCTGATGATTTTTCAGTAACTCTTGAATACATTA 1527
QY 200 Leu-----ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrlAspLeuAla 218
DB 1528 AGTCAGATGAAATGAGATCCAGCAACAAGCTACTAATATCAGAGATATAGAA----- 1581
QY 219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrlIleThrIlePhePheAsn 238
DB 1581 ----- 1581
```

```
QY 239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValAlaPheAspLeu 258
DB 1582 -----AGTGAATTGAACTCATCTTGGTGTGACTGATGATATCTG 1620
QY 259 IleLeuGluLeuTyrlAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
DB 1621 GTATTA-----GACTATGAACTACGACACAGAAAGAAAGAAATGGA 1665
QY 279 PheLysLeuLysSerAsnAspAsnGluLysArgLeuGlnValValLysLeuLeuAlaLys 298
DB 1666 TTGAATTTAAAGAAAGAAATGATTTGATGAAATTTGAGCTCTAGAAAGAAAGAACTAAA 1725
QY 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrlGlnCysTyrl 318
DB 1726 -----AAAGATCAAGACATGCA----- 1743
QY 319 LeuGlyLArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
DB 1743 ----- 1743
QY 339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrlLeuLysValArg 358
DB 1744 -----CTAATTCATGAATTTGCAACTTAAGCAATTTACTTAAGCATCGAAGATATAT 1797
QY 359 SerHisAspProGluGlnAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
DB 1798 AATCAAGATCTTGGAGATGAACTC-----AGTTCA 1827
QY 379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnDheValArgGluArgThrLeu 398
DB 1828 AAAGTACAGCTGCTTACAGAAAGAAAGACAGACAGCTTAAGAAAGCTACAGAAATCATGAC 1887
QY 399 AspLysArgTrpArgValArgLysGluAlaMetCetLysLeuGlnIleTyrlLysLys 418
DB 1888 TCTCAAAAG-----CTAGAAATATATATAAATGGACTGTGCA----- 1923
QY 419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
DB 1924 TACTCATTTGGAACATTTGAAGAACCCAAACAAATTAAGACAGACTGTTGATGCTGAA 1983
QY 439 LysLeuLeuHisIleTyrlTyrlGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
DB 1984 ACTGTA-----GCCCTTGATGCCAAGAGA 2007
QY 459 PheAlaGlnTyrlMetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyrl 478
DB 2008 GAATCAGCCCTTTCTTGAAGTGAATCTGAGTGAAGAGAAAGAAAGAACTTGCA 2067
QY 479 TyrlLeuTyrlAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
DB 2068 ACTCATATACAAAGAAATGGA-----AATGATATTCAGTTAT 2106
QY 499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
DB 2107 CAAGCCAAATTTGAGCGCAAAAGAAATGCAATGATCTGCGAAGAA----- 2157
QY 519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg----- 534
DB 2158 -----TTACATCTGCTTTTAATGAGATTAACAAACCTCACCTCTTATATAGATGCC 2208
QY 535 -----AsnLeuProAspProGlyLysLysAlaGlnAspPheMet 546
DB 2209 AAAGTTCCAAAAGATTTGCTGTGTAATTTGGAATTTGGAAGAAAGATTTACTGATCTTCAG 2268
QY 547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
DB 2269 AAAGAACTAATTAAGAAAGATTTGAAGAAAGAAAGCTTTGGCGGAAGAGTC---ATTG 2325
QY 567 ValSerProThrCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLys 586
DB 2326 CTTTCA-----GAATTGAAATCTTTACCTTCTGAGATGAAAGGCTGAGAAAGAG 2376
```

QY	587	leuGlYAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuGlu	606
Db	2377	ATA-----CAAGACAAATCTAGAGAGCTCCATATTATTAATACATGACAAAAAGAT	2424
QY	607	ArgIleAlaProValHisIleAspRhcGluSer---IleSerAlaLeuIleLysGlnVal	625
Db	2425	AAATGTGTTTCTGAAGTACGTTCATTAAGAGAGATAGAGTTCAAGGTTTACTTGAAAGAAAT	2484
QY	626	AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla	645
Db	2485	GGGAAAAACAAAGATGACCTTAGCAACTACACAGTCGATTTTAAAGACATGATCAAGAA	2544
QY	646	Ile-----	646
Db	2545	TTCCAAATTTCCAAAMCCCTTCATATGAGCTTTGAGCAAAAGTATATAGATGCTCTAG	2604
QY	647	-----ATGAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer	663
Db	2605	GAGAAATGACAGAAATGATTCAGAAATTAATCTTAATCTCTAA-----	2644
QY	664	PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp-----	680
Db	2647	-----GAAGCCCAAAATTTGATTTCGAGTTGGTGCTTTGAGACCGAGCTTCTTAC	2700
QY	681	-----AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr	695
Db	2701	AAGACCCAGAACTTCAGAGAGAAACACGCTGAGTTCAGAAAGAGCTA-----AATGAG	2754
QY	696	GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis	715
Db	2755	ATGCAACAGCTGAGAAACAATTAGAAATATGAGATTTCTCCGCTGCAAACTGTAGAAAG	2814
QY	716	HisLysSer-----	718
Db	2815	GAGAAACACATGATTACTGACAAACTGCAGCAAACTTAGAGAAAGTAAAACTTTAACT	2874
QY	719	--LysLysGlyProProArgGlnAlaLysTyralaIleHisCysIleHisAlaIlePhe	737
Db	2875	CAAGAAAAAGATGATCTTAACAACACTCCAGAAAGCTTGCAAAAT-----	2919
QY	738	SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro	757
Db	2920	--GAGAGGACCAACTCAAAAGATGATTTTACGATCTGTTAACATCAAAATATGATACT	2976
QY	758	SerAsn-----LeuGlnIleLeuIleThrProLeuValThrIleGlyHis	772
Db	2977	CAAGACACATTAACGAATGCTCTTGAGTCTGTGAACAACACTCAAGAAACATTTATATCA	3036
QY	773	IleAlaLeuLeuAlaProAspGlnPheAlaIleAlaProThrLysSerThrValAlaThrPhe	792
Db	3037	CTA-----AAATCGAAATTTCTGAGGAA	3064
QY	793	IleValLysAspLeuMetAsnAspArgLeuProGlyLysLysThrLysLeuTrp	812
Db	3061	GTATTCAGGAATTTGCATATG-----	3081
QY	813	ValProAspArgGluIleValSerProGluThrMetValLysIleGlnAlaIleLysMetMet	832
Db	3082	-----GAGGAAAAATACAGAGGAAACTAAAGATGAAATTTCCAG-----	3117
QY	833	ValArgThrPheLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg	852
Db	3118	---CAAAAGATGTTGGCATGATTAATAAAACAGAGTTTGGAAAGCTAAAAATATACC---CA	3171
QY	853	LeuLeuThrThrIleLeuHisSerAspGluLysPheLeuThrGluGlnGlyLysIleSerLys	872
Db	3172	ACACTACTCTGCAGATGTTTAAGGATTAATGAGTAATTTGACAAACAAAGGAAAGATTTTCT	3231
QY	873	-----ProAspMetSerArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAla	890
Db	3232	TTAATACAGGAGAAAAATGAACCTCCAAACAAATGTTAGAGAGTGTATTATGCGAAAAAGGA	3291
QY	891	Gln-----GluProCysTyrrHisGluIleIleThrLeuGluGlnGlyr	904

Db	3292	CAATTGAAGACGTACGACCTAAAGGAAATAATTGAATGACCATTGAAACCAGGAACAATTA	3351
Qy	905	GlnLeuCysAlaLeuAlaIleasnaspGluCysTyrGlnValArgGlnValAlaGln	924
Db	3332	AGACTT-----CTTGGGATGACTTAAAAAGCAACAGACTTATCTCCACA	3399
Qy	925	LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTyrMetAlaIleCysAla	944
Db	3400	GAAAGAACCAATGCCATTAAGAAA-----GAAGGAGAGCTTTGTAGAGCTGTGC	3450
Qy	945	LeuCysAla-----LysAspProValLysGlnArgAlaGlnAlaArgGlnLeu	962
Db	3451	AGACTGGCAGAGATTGAAGAAAACCTAAAGAAAAGACCCAGCACTCCAGAAAAACAG	3510
Qy	963	ValLysAsnIleAsnValArgArgGlnLysGlnHisAlaIleValSerGlu---	981
Db	3511	CAACAACCTTTAAATGTACAGAAAGAAAGTAGCTGAGTGCAGAAAAGATTATATAATA	3570
Qy	982	-----LysLeuLeuSerLeuLeuProGlnTyrValValProTyr	994
Db	3571	GAGAACTTAAAGAAATGAAATTAAGAAACAAGATTGACATTTGAAACATATGAAACAG	3630
Qy	995	ThrIleHisLeuLeuAlaHis--AspProAspTyrValLysValGlnAspIle--Glu	1012
Db	3631	AGCGTTGAGTTGGCTCAGAAAACCTTAATGAAATTGTGAGGAAGTAACTTATTAACAAA	3690
Qy	1013	GlnLeuLysAspValLysGlnCysLeuThrPheValIleuGlnIleLeuMetAlaLysAsn	1033
Db	3691	GAAGAAACAAGTTCTTAAGGAA-----TTACAGCAACTGTTGAAACACAG	3735
Qy	1033	GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp	1052
Db	3736	AGAGACCACTTAGAGGATATATTAAGAAATTTGAAGCTACAGCGCTACAAAACCAAGAA	3799
Qy	1053	AlaGlnGlyProAspAspAlaLysMetAsnGlnLysLeuTyrThrValCysAspVala	1072
Db	3796	GAACATAAAATTCCTCATATTCCACTTAAAGAACACCAAGAAACTATTGATGAAGTAAGA	3855
Qy	1073	MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlnSerProLysAspProVal	1099
Db	3856	AGAAACCTGA-----TCTGAGAAAGCA-----	3876
Qy	1093	LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu	1112
Db	3877	-----GCTCAATATATATTAATTACTAGAGACTTAGAAAATCCCATACCAA-----TTA	3922
Qy	1113	ProProGlnMetLysSerPhePheThrProGlnLysProLysThrThrAsnValLeuGly	1132
Db	3925	CAAGAAAGAGATCCCA-----GTGCTTCAT	3948
Qy	1133	AlaValAsnLysProLeuSerSerAlaLysGlnSerGlnThrLysSer-----Ser	1150
Db	3949	GAGGAACAAGAGTTACTCCCTTAATTCGAAAAAGTCAGCTGAGACTCAGAAAACAATGAT	4008
Qy	1151	ArgMetGlnThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle	1170
Db	4009	GAACGTGGAGTTATTAACGAAACAGTCCCAACCAAGCACTCAACACACTGGCAAGATTA	4068
Qy	1171	LysGlyArgLeuAspSerSerGlnMetLysPheHisSerGlu--AsnGluAspTyrThrMet	1189
Db	4069	-----GAAATGAAAGGCTCAGCTGGAATGAAAATTTTCACAAA	4107
Qy	1190	SerSerProLeuProGlnLysLysSerAspLysArgAspAspSerAspLeuValArgSer	1209
Db	4108	AGTCAGAGAAGATATAAATCTCTTAACCAAGGAAAGAGCAACCTTAAAAAGATAAAGAA	4167
Qy	1210	GluLeuGlnLysProArgGlyArgLysLysThrProValThrGlu-----	1224
Db	4168	GCCCTTGAGTTAAACATGACACCTGAAAACAACATATATTAGAGAAACTTTGGCTAAATTC	4227
Qy	1225	-----GlnGlnGlnLysLeuGlnMet-----AspAspLeu	1234

OY	283	berasn-----	aspsnslgluIatrgleuglnvalVallys	294
Dd	1931	GCCACGGGTATCACTAGACAGAGACTGTTGGACTTGGACAACACCATGCTCACCCAG		1990
OY	295	LeuLeuAlaLysMetPheGlYAlaLysAspSerGIuLeuAlaSerGlnAsnLysProIeu	N	314
Dd	1991	CTTCTCCTCAGTGAC-----	GCACACATGGTTCTG	2020
OY	315	TrrGLysTrYrLeuGIaYarGPhenAsnPIlenHisValProIIeaYargLeuClcysVal		334
Dd	2021	TGGTTCCTGCCACTCGGGGCTT-----	CTGCCCGAAGGCTTGAATGATGCT	2065
OY	335	LysPheaLaserHisLysCysLeuMetAsnHisProAspLeuAlaLysAspLeuThGIunYr		354
Dd	2066	CTGCACCTGTCAAGACTACGCTACAACTACCCTGTAATTCAAGAATACTATACAAAGTGT		2125
OY	335	LeuLysValArgserHisAspProGIuAlaIleaYargHisAspValIIeValSerIle		374
Dd	2126	CGCGAAGAAGTGTGAGCCGCTCACCGACGAGGAAGTGAACACGAC-----		2170
OY	375	ValThrAlaIalLysLysAspRIleuLeuValAsnAspHisIleuLeuAsnPheValArg		394
Dd	2171	-----TGC TTGATGATGGATTGATTAAGTCTCAACTAGCA		2206
OY	395	GIuaTrgThrLeuAspLysArgTrPARgValaIrGlySGluAlaMetMetGIY-----		411
Dd	2207	GATGAA--GTTGAGAAAGTGTGTGGAG--CGGCGAAGGTGGCAGACGAAATPACCT		2260
OY	412	-----LeuAlaGInIleYr-----	-----Lys	417
Dd	2261	CCTCGTAGACATCCACGCAATTACTTGATATATATATTTATAGAACACATTACCGAAGCAAGC		2320
OY	418	LysTryAla-----	-----LeuGInSerAlaIaGlyLys---	427
Dd	2321	AAACATGCTATTACCATTTATTTCGCTGTATATTATATTCCTTCCAAATRAAACG		2380
OY	428	-----AspAlaAlaLysGInIleAlatrPIIelYsAspLysIeu		440
Dd	2381	GATACCCCCCATTTGAATCTTCCCACACTGCCCTTTCCTATTTCTTGGGGCCAATTAAAGCTA		2440
OY	441	Leu-----	-----HisIleYrTrGInAsnSerIleAspArgIeu	453
Dd	2441	GTTCAAGGATTTTGGCTACTAGATCATATATGACTATATGAAATGGTTTAgAC-----	CTT	2494
OY	454	LeuValGIuaYrGliePheAlaGln-----	-----TyMetValProHis	466
Dd	2495	CTGTTTCAACCCAGTACTGCAAAACGCTGATCTGGCAACATTCAAAAGATTAATTGAAGCT		2554
OY	467	AsnIeuGIuThrTrGInArgMeLysCysIeuYrTrYrIeuYr-----		481
Dd	2555	TTTTATGAGTCAAGGAGAGCACAAACAGGCTCTCGGTATCTTTCAGACACATGAACCAACA		2614
OY	482	-----AlaThrLeuAspLeuAsnAlaVal-----	-----LysAlaIeu	492
Dd	2615	GTCGCCAGTAGACATGAAGTTATCTCTTCACTCACTGTTCTACCTTTTAAATAGATGCATG		2674
OY	493	AsnGIuMetTrpLysCys-----	GlnAsnIeuLeuArgHisGlnValLysAspLeuIeu	510
Dd	2675	GTTTAAAGCTGGAACCTTACTGCGACAGAAATTCAAACAGAGATTAATTAAGAGAAATTATTA		2734
OY	511	-----	-----AspLeuIleLysGlnPro	516
Dd	2735	AAGACAGCTTATGAAGTTTGTACAGAGATGGCGCTTANTGAGAGATTTACTGAAGCTGCCA		2794
OY	517	LysThrAspAlaSerValLysAlaIIlePheSerLysValMetValIIleThrArgAsnIeu		536
Dd	2795	TTTTCAAACACTGAG-----		2809
OY	537	ProAspProGIuLysAlaGlnAspPheMetLysLysPheThrGlnValIleuGluAspAsp		556
Dd	2810	-----CAGGATTCGCTTAAGTAATTT-----	TTACATCCAGCT	2842
OY	557	GIuLysIleArgLysGlnIeuGlnValIleuValSerProThrCysSerCysLysGlnAla		576

Db	2843	ACCACTGTTGGAATCATGAATTCCTCTTCTAGT--TCACCAATTTCACAGCGTGC--CAATTGA	2898
Qy	577	GLUGLYCSValAtrg-----:::	581
Db	2899	TATTTCTGCTCCCTTGGAACTAAACCAAGATTCTGAAATAATATCTATGATGTATGCTGACCC	2958
Qy	582	GIUlethrlYsLysLeuGlyAsnProLysGlnProThrsanProPheLeuGlu-MetII	601
Db	2959	TCGATTTGGGGAAAGATGCTGACTCGCAAAATTATATTACACAGATAGGGAAAAACCTT	3018
Qy	601	e-----LysPheLeuLeuGluArgIIleAlaProValHisIleAspThrGI	616
Db	3019	ACCTAGACTCCAGAGAAAGATTAGCTGTGGACGCAAGCTTAAGCCTTACACCGTCGTGCATGC	3078
Qy	616	uSerIle-----SerAlaLeuIleLysGlnValas	626
Db	3079	CTCAGGTTTTCATGAAGTTTCTAGACCCCAACCGTTATCGGCATTTTCACAAAG	3130
Qy	626	nlySerIleAspGlyThrAlaAspArgLysPcLuglyValaPro-----	641
Db	3131	-AAAGCTATTAAGCTGGAACAGTGTTAACCCGATCTACGCTTCATGCAAGATGTTTATGTAA	3189
Qy	642	-----ThraSpGlnAlaIleArgAlaGlyLeuGluLeuLeuLys-----	654
Db	3190	AATTGGAGAGGTGGGGCAAGCTCATGAGCCATAAAGAGCGCTCATCTTTTAAACAGTCC	3249
Qy	655	-----ValLeuSerPheThrHisProIleSerPheHisIle	666
Db	3250	TAAACAGACAGCCATCTCTGTAGAACACTCTTGTCCACACCA--	3295
Qy	666	rAlaGluThrPheGluSerLeuLeuAlaLysLeuLysMetAspArgLysValaAlaGI	686
Db	3296	-----GAGCTCTCTGCA	3306
Qy	686	uAlaAlaLeuGlnIle---PheLysAsnThrGlySerLysIleGluLysAspPheProHI	705
Db	3307	GGCGTGTGTGGAACTCCCAATTTCAATTCATCTCCACAGAAAT--	3349
Qy	705	sIleArgSerAlaLeuProValLeuHisLysSerLysLysGlyProProAlaG1	725
Db	3350	-----TCTAGATTACTGAGATTGGTTGCTGTCAT--CCTGACCCCA	3387
Qy	725	nAlaLysTryAlaIleHisCysIleHis-----AlaIlePh	737
Db	3388	GCCATTTCAGAGTGTGGAGTTTATTCACAAAGATCCACAAAGATCTCTTGTGTGCTGCT	3447
Qy	737	eSerSerLysLysLthrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsprr	757
Db	3448	GTCAGAGTCTGTTACCAATTAAGTTCACAGTTTAAAGGCGACATGACATATCTCCAGGCC	3507
Qy	757	oSerAsnLeuGlnHisLeuIleThrProLeuValThr-----IleGluHisIleAlaLe	775
Db	3508	TTCAGAGTGTCTTACTGTAGACTCTCTCATGATTAAGAAAGCTTAATCTTTGGCTGT	3567
Qy	775	uLeuAlaArgAspGlnPheAlaAlaArgProThrLysSerThrValAlaIleThrIleValu	795
Db	3568	GTCAGCCACAG-----TCTTCTGAGATTGGCCAGTTTACTCTCTCC	3606
Qy	795	sAspLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeu-----	811
Db	3607	ATGCATGCTTATGCTCT-----GGTTTTGCAAAACACCTTTACACATCTCTCTCC	3654
Qy	812	-ThrValProAspGlnLysValSerProGluThrMetValLysIIleGlnAlaIleLysMe	831
Db	3655	TTTTCTGACSTCGAAGATCTCTCACTCTCCGCTTTCAGAGTTAAAGAAACAAAGAGATTCAT	3714
Qy	831	tMetValArgThrLeuLeuGluMetLysAsnAsnHisSerLeuSerGluYthr-----	848
Db	3715	CATGGAAAGAA-----GGCATGATTAACACASTGAGATAGAGASTACAAATGACCG	3765
Qy	849	-SerThrLeuArgLeuLeuThrThrThrIleLeuHisSerAspGly-----	862

Db 3766 AATATACAAAAGCGTTTGTAGACACATCTTTCCATAAATGTGACCTTCAGACAGAACTGA 3825
 QY 863 -----Aspleu----- 864
 Db 3826 GTGGATGAAGACCACTGATAGAAATCATATTTCTCTGTGATGTCTTCGCAAAAGGCC 3885
 QY 865 -----ThrglucInglYlYsIleSerLysProasphe 875
 Db 3886 TCAGAAAGTGTGGCAGAGTCACTGGCTACCATTCAGAGAGGCTGAGAAAGCATGATGT 3945
 QY 875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysIy 885
 Db 3946 GAGCAAGAAGACAGACAGACTCCACAGCTCAGACAGACCTCCTTAGAG-----TA 3999
 QY 895 rHsgIuIleIleThrLeuGluGlnIlyrGluLeuCysAlaLeuAlaIle----- 911
 Db 4000 TCATGCGACCACTACCCAGAAAGACTTGAAGG-TCTGTGTTTGTGTCTCCCAAGCCAG 4058
 QY 912 -----AsnAspGluCysTyrGlnIy 918
 Db 4059 CATCTTCCTCCACTGAATACTAATAATTCACATCTACAAACAGAGAGG-ATAATGAT 4117
 QY 918 IArgGlnValPheAlaGlnLys-----LeuHis----- 927
 Db 4118 AAGATGCGTTTAAAGTCAAGAAAGGTACTCCTTCACCCGTGAAGAAACAAATAGCGACGG 4177
 QY 927 ----- 927
 Db 4178 GACGTGCAGTGAAGACATTTTCAGAACTGAGTCGCTTAGACCCCTGTGAAGAGCTGA 4237
 QY 928 -----LysGluLeuSerArgLeuArgLeuProleuG 938
 Db 4238 GCTTCTTTGGTGTGTGTCAGTCTGTGAAGGGAAACCTCCACTTCAAACTCCAGAGAG 4297
 QY 938 u-----TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg----- 954
 Db 4298 TCAGTTCGTGATGGAATCGTCGCTATTGAGAGCCGAACCTCATCTTACAGAGACAC 4357
 QY 954 ----- 954
 Db 4358 AAGAGTCTGTGCGCCAAACAGCGTTGCAAGATGTTGAAGCTCTGGGTCCACAGCTCCAA 4417
 QY 955 ----ArgAlaHisAlaArgGlnCysLeuValLysAsn-----11 966
 Db 4418 TGCCCGGTTAACTCTGAAGCAGCCTCGGCGAAAACTAAATTAACCTTAAAGAGAT 4477
 QY 966 eAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlnLysLeuLeuSerLe 986
 Db 4478 GAATAGAAAGCTCATGTACCAAGGAGAAAGCTGTGTTTACAGAAAGAAAGCCCTCGAAT 4537
 QY 986 uLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
 Db 4538 TCGCGTGTCTCTGTATACT-CAGAGATTCATCTAATTTGSA-----TGTGAATATCT 4590
 QY 1006 lLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTyrPheValLeuG 1026
 Db 4591 TGAAGTTCAAAATTCGAAAGAG-----GAGGCCAAGAAATCTTTCATTGATGAGTTGTA 4644
 QY 1026 uIleLeuMetAlaLysAsnGlnLysAsnSerHisAlaPheIleArgLysMetValGlnL 1046
 Db 4645 TCCCTTAGGGGCGAGAAACTTGATATATCTCACTACTATTGAGCAGACAGTTTGTGTA 4704
 QY 1046 nIleLysGlnThrLysAspAlaGlnLysProAspAspAlaLysMetAsnGlnLysLeuT 1066
 Db 4705 CTTCGCTGATGACAAAGACTCTGCTGATGTATGTGCTGAGAGTGAAGCGGAGACTTTT 4764
 QY 1066 rThrValCysAspValAlaMetAsnIleLeuMetSerLysSerThrThrTyrSerLeuG 1086
 Db 4765 T-----GTGGCCAGAGCAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 4797
 QY 1086 uSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSe 1106
 Db 4798 AGGTGAAGAAGA-----GAAGCTGAGGCAAGCGACATC 4830

QY 1106 rAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysPro-- 1125
 Db 4831 TGCAGACCTAATATTTTACCAGAAATCAGCAAGAA-----AAACCTGT 4875
 QY 1126 -----LysThrThrAsnValLeuGlnAlaValAsnLys 1136
 Db 4876 GTGCTACAGAGAAACCCCATATATCAGAGCGCGCTTACAGATTTGCCATCTGCTGACT-- 4933
 QY 1136 sProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGlnThrValSe 1156
 Db 4934 -----GCTAACCAAGATCCACAAAGGTGAGACTTTA-- 4966
 QY 1156 rAsnAlaSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSe 1176
 Db 4967 -----CCGTAATGTGCTTGAACCGGTTAAAGTGCATTCACAA 5004
 QY 1176 rSerGluMetAspHisSerGlnAsnGlnAspTyrThrMetSerSerProleuProGlyL 1196
 Db 5005 AATCTGTGGATGTAATTAAAGACACAGAAAGTAAGAGCAACTCCCGTGCAGACAG 5064
 QY 1196 sLysSerAspLysArgAspAspSerAspLeuValArgSerGlnLeuGlnLysProArg 1216
 Db 5065 TGAGCGTGTGATGAGACGAGCAGCACTGATGATCTCAAAGCTGCACATTCCTCAGAGCT 5124
 QY 1216 yArgLysLysThrProValThrGlnGlnGlnLysLeu-----G 1230
 Db 5125 GACAACTCTACACCGCAAGACTGTTAAGAACACCGTGCAGAGACTGTAAATACAGCA 5184
 QY 1230 yMetAspAspLeuThr----- 1236
 Db 5185 GAGTGATGACATGCTTCTTCTAGACTCTCACAAGAGCAGATGCCCTAAGCTGAA 5244
 QY 1236 sLeuValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAl 1256
 Db 5245 TGTCACTAGACAAAGACCTTTCAGACTGTGCCACTCTTAAGAAAGACTAGAAATAT 5304
 QY 1256 aSerGlu-----SerAspGlnGlnIlyrProGlu-- 1266
 Db 5305 TAAAGAACTCCGAGACTCTTTCGAAGAGACCTGTGTGACCTAAAGTAGACACTGAGAA 5364
 QY 1267 -----GluLysArgLeuLysGlnAspIle----- 1274
 Db 5365 CCAACTGACAGCTCAAAATCTCCCGCTCTAGAGAAAGAAAGAGAGAGCTTACGCCA 5424
 QY 1275 -----LeuGlnAsnGlnAspGlu-----GlnAsnSerProPr 1285
 Db 5425 AGGCACACTGCCAAGTTCTGTGTGTGTGGAGCCGGAGCGGAGCTCAGAGGTACGCCGG 5484
 QY 1285 oLysLysGlyLysArgGlyArgProProLysProLeuGlyLysGlyTyrPro---LysG 1304
 Db 5485 AAGACTGAGGCTGAGAAAGCAGCAGCCGAGCGAGAGAGTGAAGAAATCTCTGTAGAAC 5544
 QY 1304 uGluProThrMetLysThrSerLysLysGlySerLysLys----- 1318
 Db 5545 AAAAGTCAGAGCTTTCATCTGTAGAAAGGAAACCCCTTAAGAGACTTAAAGACTGTGTA 5604
 QY 1319 -----SerGlyProProAla----- 1323
 Db 5605 AAATGGCAAAAGATATGAATTTCTAGATGATCTCAAAAGGAGTGAAGCAGCAATCATGA 5664
 QY 1324 -----ProGluGlnGluGlnGluArg 1332
 Db 5665 CGGAGCTGCACAGACTGAGAAATCCAAATTTAGAAATATCTACAAATATGAGATATA 5724
 QY 1332 gGlnSerGlyAsnThrGlnGlnLysSerLysGlnHisArgValSerArgArgAl 1352
 Db 5725 ACAAGATGAACACAGTACAGCAGCAACCGCTTAAACCAAAAGAGGTGACAGAGAGAGA 5784
 QY 1352 aGlnGlnArgAlaGluSerProGluSerSerAlaIleGlnLysThrGln----- 1368
 Db 5785 AGTTAGTGTCTCAAGTGTGACAGAAAGCAAGACTTGAAGCTTCCAGTGTGCTCTTCA 5844

```

Qy      1369 -----SerThrProGlnLysGlyArgProSerLys 1380
Db      5845 GACAGACTGCATGTACTGCCACCCCTTAGAAA---CCTGGTAGACCCAGAGAG 5896

RESULT 14
US-09-882-227-623
/ Sequence 623, Application US/09882227
/ Publication No. US20030158396A1
/ GENERAL INFORMATION:
/ APPLICANT: Kleantous, Harold
/ APPLICANT: Al-Garawi, Amal
/ APPLICANT: Miller, Charles
/ APPLICANT: Tomb, Jean-Francois
/ APPLICANT: Oomen, Raymond P.
/ TITLE OF INVENTION: Identification of Polynucleotides
/ TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
/ FILE OF INVENTION: Genome
/ FILE REFERENCE: 06132/047002
/ CURRENT APPLICATION NUMBER: US/09/882,227
/ CURRENT FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 08/902,615
/ PRIOR FILING DATE: 1997-07-29
/ NUMBER OF SEQ ID NOS: 638
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 623
/ LENGTH: 5334
/ TYPE: DNA
/ ORGANISM: Helicobacter pylori
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (22) .. (5250)
/ US-09-882-227-623

Alignment Scores:
Pred. No.:      1.25e-09      Length:      5334
Score:          225.00      Matches:      285
Percent Similarity: 32.49%      Conservative: 217
Best Local Similarity: 18.45%      Mismatches: 523
Query Match:    3.13%      Indels:      520
DB:             12      Gaps:      71

US-09-512-581b-2 (1-1391) x US-09-882-227-623 (1-5334)
Qy      4 SerLysThrArgThrAsn---AspGlyLysIleThrTyProProGlyValLysGluIle 22
Db      955 ACAGAACTCAACCAATTTTGATGACAGACAAGTA-----GAAACAATA 999
Qy      23 SerAspLysIleSerLysGluGluMetValArgArgLysLys----- 36
Db      1000 ACTGACGACTCTACGATCAAGATGATATCAAAAGAACAAAATAATATATTGTGT 1059
Qy      37 -----MetValValLysThr 41
Db      1060 GGCATTGTAGTCGCTGTTCTTATCGTGAATATTTTATTTCTAGAACATTTTTCACCTAC 1119
Qy      42 PheMetAspMetAspGlnAspSer-----GluGluGluLysGluLeuTyLeuAsn 58
Db      1120 TTCAATGCTTTGGAAGATAAAGCTCTGCTTTTACCAAGACAGCAATCTTTATGTCAT 1179
Qy      59 LeuAlaLeuHisLeuAlaSerAspPhe-----PheLeuLysHisProGlyLysAspVal 76
Db      1180 GATGAATCCAAATATAGGCAAGATATACGATGCTGAAAGAACGGAATGAAGAAAAGGC 1239
Qy      77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyAlaProGluAlaPro 96
Db      1240 AATATGATCGAT----- 1251
Qy      97 TyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
Db      1252 -----AACAATCTTTTCTTC----- 1266
Qy      117 LeuGlnAspThrLysSerProGlnPheAsnArgTyPheTyLeuLeuGluAsnIleAla 136

```

```

Db      1267 -----AATGACGATCCCAATAGAACCTTATACAACTATTTC-----AATATGCA 1311
Qy      137 TrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGln 156
Db      1312 -----GAAATGAGAGACAAAACCCGTTG----- 1335
Qy      157 LeuTyArgThrLeuPheSerValIleAsnAsnGlnHisAsnGlnLysValHisMetHis 176
Db      1336 -----AAGACCTTTTATGATATGTATTACTTAATGTCGCACTATGAAACAATGTTTGAAG 1389
Qy      177 MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 196
Db      1390 CTTATC----- 1395
Qy      197 AspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyAsp 216
Db      1396 -----AAAGCAAAAACCTTCAAGATCAG----- 1419
Qy      217 LeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyIleThrThrPhe 236
Db      1420 -----ATGAAAAGACTCTAGAGCT----- 1440
Qy      237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPhe 256
Db      1441 TATACGACTGCATCAAAAATGCCAAACT----- 1470
Qy      257 AspLeuIleLeuGluLeuTyAsnIleAspSerHisLeuLeuSerValLeuProGln 276
Db      1470 ----- 1470
Qy      277 LeuGluPheLysLeuLysSerAsnAspAsnGluArgLeuGlnValValLysLeuLeu 296
Db      1471 -----GAAGAGAAAGATCAACTGTTTACATTTTATTC 1503
Qy      297 AlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyProGln 316
Db      1504 -----AAAGATGAAAACCTTA-----AAAAAAGCTTA----- 1530
Qy      317 CysTyLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPhe 336
Db      1531 -----CTGACCCAAACAAAAGTTCAAGTGGCGCTAGATGTTGTTGAAAAC 1575
Qy      337 Ala-----LeuMetAsnHisPro----- 345
Db      1576 GCTAAACCGATGAAAGAACGAAACGAGCGCTTAACATCAATAAGACCGTGAAGTTTGA 1635
Qy      346 -----AspLeuAlaLysAspLeuThrGluTyLeuLysValArg 358
Db      1636 GAGAAATTCGCTAAGCAATTAAGACTCTCAAAAAGAGCTTCAAGAGTAT----- 1683
Qy      359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
Db      1684 -----AAGATTGTATCAAAAACGCCAAACACAGAACT 1716
Qy      379 LysLysAspIleLeuLeu-----ValAsnAspHisLeuAsnPheValArgGluArg 396
Db      1717 GAGAAAACCAATGCTTGAAGAGCTTGCTTAAGAAGCATATAGAGATGTGAACACGAA 1776
Qy      397 ThrLeuAsp-----LysArgThrArgValArgLysGluAlaMetMetGly 411
Db      1777 GCGTAGATTTGTTGAAGACGCTAAACCGATGAAGAACGAAACGAGTGTGTAATAAT 1836
Qy      412 LeuAlaGlnIleTyTyLysLysTyAlaLeuGln----- 422
Db      1837 ATTCGCCCAAGACTTGCAAAAAGAACTATTACTGTATAGACGTCACAGCTTACAGAGAT 1896
Qy      423 -----SerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
Db      1897 TCGCATCAAAAAGCTAGAAATGAAGAAAGACAAACAGATGC-----GAGCAATTG 1947
Qy      441 LeuHisIleTyTyGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
Db      1948 CTCACGCTGAAAGGAGAGAAAAGTTAGAACACACAGCTTCA----- 1989

```

QY 461 GlnTyrMetValProHisAsnLeuGluThr---ThrGluArgMetLysCysLeuTyrTyr 479
 ||| :||| ||||| ||||| |||||
 Db 1990 -----GATGTTTGAAGAAACCGTAAGACGATGACAGCAAGAAAGAGTGTTCGAAGAT 2043
 QY 480 LeuTyrAlaThrLeu-----AspLeuAsnAlaValLysAlaLeuAsn----- 493
 ||| ||| ||||| ||| :||| |||
 Db 2044 CTCCTTAAAGACTTACAAAGCGATATCTAGCCAAAGAGCCGTAAGCTTATAAGAC 2103
 QY 494 -----GluMetTrpLysCysGlnAsnLeu----- 502
 ||| ||| :||| ||| ||||| |||||
 Db 2104 TCGGTATCTCAAGCCAAACCGAAGCTGAGAAAAAAAGAAATGCGGAATTTACTCACCCCT 2163
 QY 503 -----ArgHisGlnValLysAspLeuAspLeu 512
 ||| ||| ||||| ||||| |||||
 Db 2164 GAAGCGAAAAACTTTAGAGAAGAGAGCAAGAGAGCGTTAAGCTTATTTCGATTCG 2223
 QY 513 IleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIle 532
 :||| ||||| ||||| ||||| |||||
 Db 2224 GTATCTCAAGCCAAACCGAAGCTGAGAAAAAAGAAATGCGGAATTTACTCACCCCT 2277
 QY 533 ThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnVal 552
 ||||| ||| ||| :||| ||||| |||||
 Db 2278 -----CTGAAGCGAAAAAAAGTTAGAGAAGCTTAAGAAAAAGCGTTAAAGCT 2325
 QY 553 LeuGluAsp-----AspGlnLysIleArgLysGlnLeuGluValLeu 566
 ||| :||| ||| :||| ||| :||| |||
 Db 2326 TACTGTGATTCGCGATCAAGAGCTAGGAATGAAAAAGAAAAAGAAATGCGAGAAATTTG 2385
 QY 567 ValSerProThrCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLys 586
 :||| ||| ||| ||| :||| |||
 Db 2386 CTCACCCCTGAAGCGAAA---AAACTTTAGAGCAACAGCACTAATGTTTTCGAAGAAC 2442
 QY 587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPheLeuLeuGlu 606
 ||||| ||| ||| :||| ||| ||| |||
 Db 2443 GCTAAAAACGATTAAGAACAAAAAGTGTTCGAAGATCTCCCTTAAGACTTCGACAAA 2502
 QY 607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
 :||| ||||| ||| :||| ||| :||| |||
 Db 2503 AAGGTT-----TTAGCTAAAGAAAGCGTTAAAGCTTACTTGATTCGATTCG 2550
 QY 627 LysSerIleAspGlyThrAlaAspAspGluAspGluValAlaProThrArgGlnAlaIle 646
 :||| ||| ||| ||| :||| |||
 Db 2551 CAAGCCAAACCTGAAGCTGAGAAAAAGAAATGCGAAATTTACTCCAGCCCTGAAGCG--- 2607
 QY 647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
 ||| ||||| ||| ||| :||| |||
 Db 2608 AGAAACTTTTAGAAGAAAGCTAAA----- 2631
 QY 667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
 :||| ||||| ||| ||| :||| |||
 Db 2632 ---AAAAGCGTTAAGCGCTTATTTGGATTGC-----GTAATCTCAA 2667
 QY 687 AlaAlaLeuGlnIlePheLysAsnThrClySerLysIleGluGluAspPheProHisIle 706
 ||| ||| ||||| ||| ||| :||| |||
 Db 2668 GCGAAAACTAGAGCTGAGAAAAAGAAATGCGAAATTTACTCAAC-----CTGAAAGCG 2721
 QY 707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysGlyProProArgGlnAla 726
 ||| ||| ||| :||| ||| :||| |||
 Db 2722 AGAAACTCTTA-----GAAGAAGCT 2742
 QY 727 LysTyrAlaIle-----HisCysIleHisAlaIlePheSerSerLysGluThr 742
 ||| :||| ||| :||| ||| :||| |||
 Db 2743 AAAGAGAGCGCTTAAAGCTTATAAGAGCTGCGTATCAAAAGCTAGGAATGAAAAAGAGAAA 2802
 QY 743 GlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerGlnLeuHis 762
 :||| ||| ||| :||| ||| :||| |||
 Db 2803 AAA-----GAATGCGAAGAAA 2817
 QY 763 LeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProArgGlnPheAla 782
 ||| :||| ||||| ||| ||| :||| |||
 Db 2818 TTACTCAGCGCT-----GAAGCGAAAAAACTTTTAGAGCAACAA----- 2856

QY 783 AlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuMetAsnAspArg 802
 ||| ||| ||| ||| :||| |||
 Db 2857 -----GTGTCATGATGTTTGAAGAAACGCTAAA 2883
 QY 803 LeuProGlyLysGlyThrThrLysLeuThrValAlaProAsp-----GluGluValSerPro 820
 ||| ||| ||| ||| :||| |||
 Db 2884 ACCGAAACGTATAAA-----AAAAGCTGTCAAAAGATCTCCCTAAAGACTTCGACGAAA 2937
 QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeuLeuGluMetLys 840
 :||| ||| ||| ||| :||| |||
 Db 2938 AAGGTTTACCTAAA---GAGAGCGTTAAAGCTTATTTGGACTGCGATGAAAGCTAGG 2994
 QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
 ||| ||| ||| :||| ||| :||| |||
 Db 2995 AATGAAGAAAGAGAAAAA----- 3012
 QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuAlaGluLeu 880
 ||| ||| ||| ||| :||| |||
 Db 3013 -----GAATGCGAAGAAATTTCTCAACCCCTGACAGCGAAAAAAGCTTTAGAA 3057
 QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThr 900
 ||| ||| ||| ||| :||| |||
 Db 3058 GAAGCCAAAGAGACTTT-----AAAGCTTATAAGACTGCTCTCTCAAGCTAGAAAT 3111
 QY 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
 ||||| :||| ||| :||| ||| :||| |||
 Db 3112 GAAGAAAGAAAGAGAGCTTCGAGAAAGCTCAAGCGCTGAA-----GCGAGAGAAA 3162
 QY 921 ValPheAlaGlnLysLeuHisLysGlyLeu-----SerArgLeu 933
 :||| ||| ||| ||| :||| |||
 Db 3163 CTCTTAGAGCAAGAAAGTTAAGAAAGCACTTAAGGCTTATTTGGACTGCGATGAAAGCT 3222
 QY 934 ArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
 ||| ||| ||| ||| :||| |||
 Db 3223 AGGATGAAAAAGAGAAAAAGAAATGCGAAGAAATTTCTCAAGCGCTGAGAGCAAGAAATTT 3282
 QY 954 ArgArgAlaHisAlaArgGlnCysLeuValLys-----AsnIleAsnValArgArgGlu 971
 ||||| ||| ||| ||| :||| |||
 Db 3283 TTAGCGCAAGCAAGCTGTAATTTGTTGGAAAAAGCTCGAAATGAGCAAGAAAGAAAGCA 3342
 QY 972 TyrLeuLysGlnHisAlaIleValSerGluLysLeuLeuSerLeuLeuProGlyTyrVal 991
 ||||| ||||| ||||| ||||| |||||
 Db 3343 TGCTTAAAAAT-----CTCCTTAAGACTTA 3369
 QY 992 ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle 1011
 :||| ||||| ||||| ||||| |||||
 Db 3370 CAGGAA-----AATATTTTAGCTAAA----- 3390
 QY 1012 GluGlnLeuLysAspValLysGluCysLeuThrPheValLeuGluIleLeuMetAlaLys 1031
 ||| ||||| ||||| ||||| |||||
 Db 3391 GAGAGCTTTAAAGCTTATAAGAACTGCTC-----TCTCAAGCTAGA 3432
 QY 1032 AsnGlnAsnSerHisAlaPhe-----IleArgLysMetVal 1044
 ||||| ||| ||| ||| :||| |||
 Db 3433 AATGAAGAAAGAGAGAGCTTCGAGAAAGCTCAAGCGCTGAGAGAGAAAGCTTTA 3492
 QY 1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
 ||| ||| ||| :||| ||| :||| |||
 Db 3493 GAG-----CAAGAACTTAAGAAAGCGTTAAG 3519
 QY 1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
 ||| ||| ||| ||| :||| |||
 Db 3520 GCTTATTTGAGCTGCGATGACAGAGCTAGAAATGAAAAAGAGAAAGAAATGCGAGAAA 3579
 QY 1085 LeuGluSerProLysAspProValLeuProAlaArgPheThrThrGlnProAspLysAsn 1104
 ||| :||| ||||| ||| ||| :||| |||
 Db 3580 TTACTCAGCGCTGAGAGGAGAAATTTTTAGCGAAAGAACTCCAAACAAAGATTAAGCG 3639
 QY 1105 PheSerAsn---ThrLysAsnTyrLeuPro----- 1113
 :||| ||||| ||||| ||| :||| |||
 Db 3640 ATCAAGAAATTTGCTTGAAGAAAGCGCGATCTTAAGAGAGAGAGCGGCTATCATGAAGTGTTCG 3699
 QY 1114 -----ProGluMetLysSerPhePheThrProGlyLysProLysThrThr 1128


```

Db      3700 GATGTTTGAGCGATGAGAGAAAGCTCAATACCTGCAAGAGCTAGAGAAAAAGCGTGT 3759
      1129 AsnValLeuGlyAlaValAlaValProLeuSerSerAlaGlyLysLeuIntrLys 1148
      3760 GCGGATGTTGGCTATGAGCTTAAGACGATGAAGAAAAAGGAGCCAAACCTTAT 3819
      1149 SerSerArgMetGlyThrValSerAlaSerSerSerSerSerProSerProGly 1168
      3820 ACCGATTGATCCAGAAATCCAAATTAAGACACCAAAACAAATCAATGAGT 3879
      1169 ArgIleLysGlyArgLeuAspSerSerGlu---MetAspHisSerGluasn-----Glu 1185
      3880 AAAACAGAAAGGTTGCATCAACAGCAACGAGTGGCTGTGATTAACCTAGATGACCTACTGAT 3939
      1186 AspTyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg----- 1201
      3940 CAAGAGCCATAGACCAATGTTTAAAGGCTTGAGCGATAGTAAAGGCGCTAATTCAT 3999
      1202 -----AspAspSerAspLeuValArgSerGluLeuGlyLysProArg 1215
      4000 GGAATTTAAACGACAGCTGATGAGTGGATCTGATTTATAGCATTAAGAAACCGTAA 4059
      1216 -----GlyArgLysLysThrProValThrGluGluGlyLysLeuGly 1230
      4060 ACCTTTGATACATGCGCGCTAAAGGTATTCATTTGTTACCAATGATTTCAAAATGGC 4119
      1231 MetAsp-----AspLeuThrLysLeuValGluGluGln 1241
      4120 GGCAGATTTGCCACTATTAAACCCCACTAATGTTGATGCGGCAAAATACCTAGCGTAAAT 4179
      1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp--- 1260
      4180 CCTATT-----TATGCTTCCATAAGCGCTGATATT 4209
      1261 GLuGlnGlnTTPProGluGluLysArgLeuLysGlyAspIleLeuGluAsnGluAspGlu 1280
      4210 GCCAGCAATACGAAACAGAAAAACCATTAAGATTAAGAAATTTAGAA----- 4257
      1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
      4258 -----GCTAAATTAAGCTAAGGCTTATGGTGCAT 4287
      1301 ThrProLysGluGlnProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
      4288 -----AAAAAGATACGATTAAGAAAAAGTAAAAATCCACACAGCAAGCTAAAGCA 4341
      1321 ProProAlaProGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
      4342 GAAACCAATTAAGATAGCAAAAGTGTCCGAGAAATGCGCAAGAAATATCAGTAAATCGCT 4401
      1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnArgAlaGlu-----Ser 1358
      4402 CTTTAAGAACAAAAAAGAAAGAGTGGGAATTTAGTAAAGAAATGTAATCCCATTTGAT 4461
      1359 ProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLys-----Gly 1374
      4462 GACAAAGAAAGAAAGCAAAACAAAGATGAACAGCCCTGTCTCAACAGGCCCTTTATAGCC 4521
      1375 ArgGlyArgProSer 1379
      4522 AAGAGTGAATCCACA 4536

```

```

RESULT 15
US-10-171-581-49
; Sequence 49, Application US/10171581
; Publication NO. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999

```

```

; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 49
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: Homo sapiens
; DATABASE INFORMATION:
; DATABASE ACCESSION NUMBER: D21262
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-49

Alignment Scores:
Pred. No.: 4,26e-08 Length: 3727
Score: 205.50 Matches: 141
Percent Similarity: 33.64% Conservative: 77
Best Local Similarity: 21.76% Mismatches: 256
Query Match: 2,86% Indels: 175
DB: Gaps: 26

US-09-512-581b-2 (1-1391) x US-10-171-581-49 (1-3727)
      858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
      115 CTACACAGC-----AGATGCCAATGCTCTTCCCTCTTAA 150
      878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnLubProCysTyrHisGlu 897
      151 ACATCATAGCTCTTCTGCTCAAGTGTCCAGAG---TCCACAGCGCAAGATTACAGAGCA 207
      898 IleIleThrLeuGlnGlnLysGlnLysCysAlaLeuAlaIleAsnAspLysCysTyrGln 917
      208 -----ATGACCAGTGG---CTAAGAAAGCTAAGAAAGGCTCATCCAGTGACA 255
      918 ValArgGlnValAlaPhe---AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg----- 934
      256 GTGAGGACAGCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 315
      935 ---LeuProLeuGlnLysMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
      316 TACCTGCCAAGCAGAG-----TCGGTCTCCCTCTGGGAAGGCTGCAGCCAAAGCATTCAG 369
      954 ArgArgAlaHisAlaArgGlnLysLeuVal-----LysAsnIle 966
      370 AGAGTAGCAGCAGTGAAGATCCAGATGATGATGATGATGATGATGATGATGATGATGATG 429
      967 AsnValArgArgGlnLysLeuLysGlnHisAlaIleValSerGluLysLeuLeuSerLeu 986
      430 CTGTCCAGAAAGGAGGATTAAGCCCAAGCCCAAGG-----CAGCCAAAGCTCTCTCTTA 483
      987 LeuProGluLysValVal-----ProTyrThrIleHisLeuLeuAla 1000
      484 AGGCCAAGAGCTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTTCTGATTT 543
      1001 HisAspProAspTyr-----ValLysValGlnAspIleLeuInLysAsp 1016
      544 AGAAG-CCAAAGATTAACACCTGTGACAGTTAAAGCTCAAGATTAAGCCCTCCCAAGCA 602
      1017 ValLysGlnCysLeuThrPheValLeuGluIleLeuMetAlaLysAsnGluAsnSer 1036
      603 GCTCGAGCAGCA-----CTAATAATATCCCAATGTTAA 635
      1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
      636 GCAGCCAGTAGCAGCAGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 695
      1057 AspAspAlaLysMetAsnGlnLysLeuLysTyrThrValCysAspValAlaMetAsnIleIle 1076
      696 AAGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 755
      1077 MetSerLysSerThrThrTyrSerLeuGlnSerProLysAspProValLeuProAlaArg 1096

```

```

Db      756 CCTAAAAAGCAAGTTGTG-----GCCAAGGCCCCAGTGAAGACAGCTACC      800
Qy      1097 Pheherhr-----
Db      801 ACCCCTACCCGGAAGAGCTTACGACGTGAGGATTCCTCCAGTACGAGAGAGAGACAA      860
Qy      1100 ---GlnProAspLysAsnPheserAsnThrLysAsnThrLeuProProGluMetLysSer      1118
Db      861 AAAAACCACCATGAATAATAAACCAAGTCCCTACAGTTCAGTCCCGG-----      908
Qy      1119 PheherhrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu      1138
Db      909 -----CCTTCTGCTCCCGCCACCAAGAGTCTCTGGGA---ACCCAGCTCTCCCAAG      956
Qy      1139 SerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAla      1158
Db      957 AAGCTGTGGAGAGACACACACCTGTGTGAAGCAGT-----GAAACACGACGTGATGAG      1010
Qy      1159 SerSerSerSerAsn-----
Db      1011 TCTGATTCAGTCTCTGAAGAAGAGAGAAACCCCACTAAGCAGTAGTCTCTAAGCA      1070
Qy      1164 -----ProSerSerProGlyArgLysGlyArgLeuAspSerSerGluMet      1179
Db      1071 ACCACTAAACCACTCCAGCAAGAAAGACAGACAGACCTCTCAGACAGCTCAGACTCT      1130
Qy      1180 AspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysSerAsp      1199
Db      1131 GACAGCTCTGAG---GATGATGAAGCTCTCTTAAAGCCAGCTGTACCACCAAGATTTCT      1187
Qy      1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys      1219
Db      1188 TCAATATAGCCAGCTGTACACCAAGTCACTGACGTGAAGCCAGCTGCAGCCCGCAAG      1247
Qy      1220 ThrProVal-----Thrglu      1224
Db      1248 CAACCTGTGGGGGTGGCCAGACCTTCTGACGAGAAAGCTGACAGCAGCTCCAGTGAG      1307
Qy      1225 GlnGlnGluLysLeuGlyMetAspAspLeuThr---LysLeuValGlnGlnLysPro      1243
Db      1308 GAAGAGAGCACTCCAGTGAAGAGAGAGAGACAAAGAGATGTGGCCACCACTAAGCCC      1367
Qy      1244 LysGlySerGlnArgSer-----ArgLysArgGlyHisThr      1255
Db      1368 AAGCGACTGCCAAGAGCTCTACTCTGCTGCCAGCAAGCAGGCTCTCAGGAGTAGAG      1427
Qy      1256 AlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspLeu      1275
Db      1428 GACAGCAGCTCTGAT-----TCAGACAGCTCC      1454
Qy      1276 GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys      1295
Db      1455 AGCAGTGAAGAGAGAGAGAGACATCTAAGTCTGCAATTAAAGAAAGCCACAGAG      1514
Qy      1296 ProLeuGlyGlyThrPro-----
Db      1515 GTACAGAGAGGTGACAGCCCTTCCAAAGCCAGCTCTGCAAGAAAGAAAGGCTGAGAGC      1574
Qy      1303 -----LysGluGluProThrMetLysThrSerLysLys      1313
Db      1575 AGCAACAGTTCTTCTGTGATGACTCCAGTGAAGAGAGAGAAAGCTCAAGGGCAAG      1634
Qy      1314 GlySerLysLysSerGlyProProAla-----
Db      1635 GGTCTCTCCAGACCAACAGCCCAAGGCCAATGGCACCTGCACTGATGCCAGAAAT      1694
Qy      1324 -----ProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu      1337
Db      1695 GGAAGAGCACTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG      1754
Qy      1338 GluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnArgAlaGlu      1357

```

```

Db      1755 GTTTCAAATCAGGTTCTATTAAAGACGGAAGCAGAGATGAGCTGCCAAGAGAGCAGAG      1814
Qy      1358 SerProGluSerSerAlaIleGlu---SerThrGlnSerThrProGlnLysGlyArgGly      1376
Db      1815 ACTCTCAAGGCCAAGAGATTAAGCTTCAGACCCCTTAACACATTTCCAAAAAGAGAGAA      1874
Qy      1377 ArgProSerLysThrProSerPro      1384
Db      1875 GGAGAAAAAAGGCGCATCATCCCA      1898

```

Search completed: September 25, 2003, 04:38:19
Job time : 1030 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:26:40 ; Search time 167 Seconds

(without alignments)
3676.431 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRNDGKITRYPGCVK.....QKGRGSPSKTPSPQPKKNV 1391

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09512581/runat_23092003.163547.5508/app.query.fasta_1.1543
-DB-Issued_Patents_NA -QFWT=fastap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptio -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09512581.ecgn1.1.85.etrnat.23092003.163547.5508 -NCPD=6 -ICPD=3
-NO_MMAR -LARGEOBERT -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

- 1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2.6/ptodata/1/ina/PCRTUS.COMB.seq:*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	349	4.9	530	3	US-09-328-111-682
2	231	3.2	8503	4	US-09-620-312D-130
3	229	3.2	8257	4	US-09-620-312D-130
4	207	2.9	6755	3	US-09-595-684B-30
5	201.5	2.8	10136	1	US-08-931-999-4
6	201.5	2.8	10136	5	US-08-353-700-2
7	197	2.7	8789	1	PCR-US95-16216-2
8	191	2.7	10320	4	US-08-328-254-5
9	190	2.6	30549	4	US-09-091-501B-9
10	189.5	2.6	4868	4	US-09-134-001C-322
11	189.5	2.6	4868	5	US-08-139-937-12
C 12	184.5	2.6	19307	3	PCR-US93-11310-12
					Sequence 10, Appl

C 13	184.5	2.6	19307	3	US-09-427-048A-10	Sequence 10, Appl
14	182.5	2.5	6306	1	US-08-195-487-3	Sequence 3, Appl
15	182.5	2.5	6306	5	PCR-US93-06160-3	Sequence 3, Appl
16	181	2.5	6605	1	US-08-769-309A-4	Sequence 4, Appl
17	181	2.5	6605	4	US-08-994-570-4	Sequence 4, Appl
18	181	2.5	6608	4	US-09-220-132-58	Sequence 58, Appl
19	180.5	2.5	13977	4	US-09-484-970B-60	Sequence 60, Appl
20	180	2.5	8351	1	US-08-198-446B-14	Sequence 14, Appl
21	180	2.5	8351	2	US-08-870-669-14	Sequence 14, Appl
22	179.5	2.5	6306	1	US-08-466-390-3	Sequence 3, Appl
23	179.5	2.5	6306	1	US-08-470-950-3	Sequence 3, Appl
24	179.5	2.5	6306	1	US-08-467-781-3	Sequence 3, Appl
25	179.5	2.5	6306	2	US-08-483-924-3	Sequence 3, Appl
26	179.5	2.5	7453	4	US-09-620-312D-248	Sequence 248, App
27	178	2.5	6152	4	US-08-973-462-1	Sequence 1, Appl
28	177	2.5	6921	4	US-09-643-597-117	Sequence 117, App
29	177	2.5	6921	4	US-09-480-884A-117	Sequence 117, App
30	177	2.5	6921	4	US-09-542-615A-117	Sequence 117, App
31	177	2.5	6921	4	US-09-606-421B-117	Sequence 117, App
32	176.5	2.5	7501	4	US-09-620-312D-249	Sequence 249, App
33	176	2.4	4716	2	US-08-290-731C-1	Sequence 1, Appl
34	175.5	2.4	6775	4	US-09-620-312D-289	Sequence 289, App
35	174.5	2.4	6519	1	US-08-588-985-1	Sequence 1, Appl
36	174.5	2.4	6519	1	US-08-971-988-1	Sequence 1, Appl
37	174	2.4	5361	3	US-08-973-462-2	Sequence 2, Appl
38	173.5	2.4	6773	4	US-09-166-350-27	Sequence 27, Appl
39	171	2.4	580073	4	US-08-545-528D-1	Sequence 21, Appl
40	170	2.4	9370	1	US-08-320-558-27	Sequence 27, Appl
41	170	2.4	9370	5	US-08-545-860D-27	Sequence 27, Appl
42	170	2.4	9370	5	PCR-US94-04496-27	Sequence 27, Appl
43	170	2.4	9391	3	US-08-320-559-25	Sequence 25, Appl
44	170	2.4	9391	3	US-08-545-860D-25	Sequence 25, Appl
45	170	2.4	9391	5	PCR-US94-04496-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-682/C
Sequence 682, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endeago, Wilison O.
APPLICANT: Steilmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adrian
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 682
LENGTH: 530
TYPE: DNA
ORGANISM: Homo sapiens
US-09-328-111-682
Alignment Scores:
Pred. No.: 1,94e-24
Score: 349.00
Percent Similarity: 98.53%
Length: 530
Matches: 67
Conservative: 0

Best Local Similarity: 98.53% Mismatches: 1
Query Match: 4.85% Indels: 0
DB: 3 Gaps: 0

US-09-512-581b-2 (1-1391) x US-09-328-111-682 (1-530)

QY 37 MetValValIysThrPheMetAspMetAspGlnAspSerGluGluValArgArgLeuTyr 56
Db 213 ATGGTTGTGAANAOTTTTATGATATGACACAGACTCTGACAGAGAAAAGACCTTAT 154
QY 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
Db 153 TTAACCTACCTTACATCTTGCTTCAGATTTTCTCAAGCATCTGATTAAGATGT 94
QY 77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
Db 93 CGCTTACTGTAGCTCTGCTGCTGATATTTTCAGATTAATGTCTCCGAAAGCTCT 34
QY 97 TyrThrSerProAspLysLeuLys 104
Db 33 TACACATCCCTGATTAACATAAG 10

RESULT 2

US-09-620-312D-130
Sequence 130, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, AIdong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784C1P2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes Version 1.0
SEQ ID NO 130
LENGTH: 8503
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (91)..(8082)
US-09-620-312D-130

Alignment Scores:

Pred. No.: 6.67e-11 Length: 8503
Score: 231.00 Matches: 285
Percent Similarity: 35.41% Conservative: 267
Best Local Similarity: 18.28% Mismatches: 541
Query Match: 3.21% Indels: 467
DB: 4 Gaps: 67

US-09-512-581b-2 (1-1391) x US-09-620-312D-130 (1-8503)

QY 17 ProGlyValIysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
Db 1087 CCTATGTATATAGGATCA-----ACTGATGAAGCTCTCCGAAAGGATATCA 1137
QY 37 MetValValIysThrPheMetAspMetAspGlnAspSerGluGlu----- 51
Db 1138 -----AAAGAAATATGATCTTAAAAAACATTAAGAGAGGTTCTTAGAGACG 1188
QY 52 -----GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db 1189 CGGGCTCAGCAATGCAAAAAAGACCAATTGGCCCACTTTGGAAGAA-----AAAGAT 1242
QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
Db 1243 TTGCTTCAGAAAGTACGAAATGAGAAATTAAGAACTTAACACGAGATGCTG 1296
QY 83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
Db 1297 -----ACCTCTCTTCC----- 1308
QY 103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
Db 1309 -----CTCAGCTTGCA-----CAGGATTAAGAGCT 1335
QY 123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAlaSerGlnIleAlaThrValLysSerTyrAsn 142
Db 1336 AAAAGAAAGAAAGAGTACTTGCTGCTGGCAAAATTAACAAATTAAGAAAGCTCAAC 1395
QY 143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
Db 1396 -----TATGAGATCAATTT 1410
QY 163 SerValIleAsnAsnGlnHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
Db 1411 AATATACCAACAAT---ATAACAACAACAACATAGCTTTCTAATATTTATATACGA 1467
QY 183 SerIle-----IleCysGluGlyAspThrValSerGlnIleLeuLeuAspThrVal 199
Db 1468 GAATATGATGATCTGCTGCTTCAGAGTGCATGATTTTCAGTAACCTGTGATACATT 1527
QY 200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
Db 1528 AGTCAGATGAAATGCAATCCAGCAACAACCTTAATTCAGGAGATATAGAAA----- 1581
QY 219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238
Db 1581 ----- 1581
QY 239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
Db 1582 -----AGTGAATTACCTCACTTGCTGCTGCTGCTGATATATCTG 1620
QY 259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
Db 1621 GTATT-----GACTATGACAACACTACGACAGCAAAAAAATAATGCAA 1665
QY 279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValAlaValLysLeuLeuAlaLys 298
Db 1666 TTGAATTAAGAAAGAAAGATGATTTGGATGAATTTGAGGCTCTAGAGAAAGAAACTAAA 1725
QY 299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyr 318
Db 1726 -----AAAGATCAAGAGATGCAA----- 1743
QY 319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
Db 1743 ----- 1743
QY 339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
Db 1744 -----CTAATTCATGAATTTTCGAACCTTAAGAAATTTAGTACGATCGAAGATATAT 1797

OY	359	SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaIa	378
Db	1798	AATCAAGATCTTCTGGAATGAATCAATC-----AGTTCA	1827
OY	379	LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu	398
Db	1828	AAAGTAGAGCTGCTTAGAGAAAGAAACACAGATTAAAGAGCTACAGAAATACATAGAC	1887
OY	399	AspLysArgTyrPArgValArgLysGluIleMetMetGlyLeuAlaGlnIleTyrLysLys	418
Db	1888	TTCTCAAAAG-----CTAGAAATATTAATAATGACTTGCTCA-----	1922
OY	419	TyrAlaLeuGlnSerAlaIleGlyLysAspAlaValLysGlnIleAlaTrpIleLysAsp	438
Db	1924	TACTCATGTGAAACCATGATGAGAACCCCAAAACAAATGAAGCAGACTCTGTTGAGCTGCA	1983
OY	439	LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArgLeuValGluArgIle	458
Db	1984	ACTGTAA-----GGCCTTGATGCCAAAGAGA	2007
OY	459	PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr	478
Db	2008	GAATCAGCGCTTCTTACAGAGTGAATATGTGAGTTGAAGGAGAAATATAAGAACTTCA	2067
OY	479	TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys	498
Db	2068	ACTCATATCAACAATATGGAA-----AATGATATTCAGTTATAT	2106
OY	499	GlnAsnLeuAlaGHisGlnValLysAspLeuAsnLeuIleLysGlnProLysThr	518
Db	2107	CAAGGCCAATTTGGAGGCAAAAGAAATGCAAGTTGATCTGGAGAAAGAA-----	2157
OY	519	AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg-----	534
Db	2158	-----TTACATCTGCTTTTATATAGATACAAACATCACTACCTCCCTTATAGATGCC	2208
OY	535	-----AsnLeuProAspProGlyLysAlaGlnAspPheMet	546
Db	2209	AAAGTTCCAAAAGATTTGCTGCTGTAAATTGTGAATTCGAAGAAAGATTTACATGATCTTCAG	2268
OY	547	LysLysPheThrArgValLeuGlnAspAspGluLysIleArgLysGlnLeuGluLeu	566
Db	2269	AAAGAACATAAATAAGAAATGGAAGAAATGCACTTGGCCGGAAGAAAGTC--ATTATTG	2325
OY	567	ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys	586
Db	2326	CTTTCA-----GAATTGAAATCTTTACCTTCTGAGATGAAAGGCTTAGGAAAGAG	2376
OY	587	LeuGlnLysnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu	606
Db	2377	ATA-----CAAGACAAATCTGGAAGAGCTCCATTAATTAATCACTCAAGAAAGAT	2424
OY	607	ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal	625
Db	2425	AAATTTGTTTTCTGAAGTAGTTCATAGAGAGAGTAGTCAAGGTTTACTTGAAGAAATTT	2484
OY	626	AsnLysSerIleAspGlyThrAlaAspAspGluLysGluValAlaProThrAspGlnAla	645
Db	2485	GGGAAACAAAGATGACCTAGCAACTACACAGTCGCAATTATTAAGACATGTATCAAGAA	2544
OY	646	Ile-----	646
Db	2545	TTCCAAATATTTCAAAACCCCTTCATATGAGACTTTGAGCAAAAGTATAGATGCTGCTTCAG	2604
OY	647	-----ArgIaGlyLeuGlnLeuLeuLysValLeuSerPheThrHisProIleSer	663
Db	2605	GAGAAATGAGAAATGAATCAAGAAATAGTTATCTCTCTAANA-----	2646
OY	664	PheHisSerAlaGluThrPheGlnSerLeuLeuLysLeuLysLysMetAsp-----	680
Db	2647	-----GAAGCCCAAAATTTGATTCAGAGTTGGGCTGCTTTGAAGACCGACCTTCTTAC	2700
OY	681	-----AspGluLysValAlaGluIleAlaLeuGlnIlePheLysAsnThr	695

Db	2701	AAGACCCAAAGACTTTCAGAGAGAAAAACAGCTAGGTTCAGAAAGAACTA----				2754
QY	696	GLYSERLSTILEGIUGLISAPRhePROHisTLeaTgSeratLaLeuProValLeuHis					715
Db	2755	ATGCAACAGCTGAAAGAAACAATTGAAATAATGAGATTCTCCGTGCAAACTGTAGAAAG					2814
QY	716	HISYSer-----					718
Db	2815	GAGAAACACGATTACTGAGAAACCTGCAGAAACTTGAAGAAAGTAAAAACTTTAACT					2874
QY	719	---LysLysGIYProProaTgGlnLaLysTyrLaIleHisCysLIEHisLaIlePhe					737
Db	2875	CAGCAAAAAAGATGATCTTAAACAACTCCAGAAAGCTTGCAAAAT-----					2919
QY	738	SerSerLysGIUThnGlnPheLaGlnLlePheGluProLeuHisLysSerLeuAsPro					757
Db	2920	---GAGAGGGACCAACTTCAAAAGGATATATACATACTGTGTTAACTGAATATGATACT					2976
QY	758	SerAsn-----LeuGlnHisLeuLleThrProLeuValThrLIEGlnHis					772
Db	2977	CAAGACAACTTACGAAATGCTCTTGAGCTCTGTGAACAACAACATCAAGAAACATTAATATCA					3036
QY	773	LlealaLeuLeuLaIleProAspGlnPheLaIleLaIleProTyrLysSerTyrLaIleThrPhe					792
Db	3037	CTA-----AAATGCAAAATTTCTGAGGAA					3060
QY	793	LleValLysAspLeuLeuMetAsnAspArgLeuProGluLysLysThnLysLeuTrp					812
Db	3061	GTTCCTCAGAAATTGTCATATG-----					3081
QY	813	ValProAspGlnGluValSerProGluThnMetValLysLIEGlnLaIleLysMetMet					832
Db	3082	---GAGGAAAAATACAGAGAAACCTAAAGATGAATTTCAG-----					3117
QY	833	ValArgTyrPLeuLysGluMetLysAsnAsnHisSerLysSerGluThnSerThnLeuArg					852
Db	3118	---CAAAAGATGGTGGCATGATGATAAAAAACAGAGATTTCGAAAGCTAAAAAATATAC---CAA					3171
QY	853	LeuLeuThnThnLIELeuHisSerAspLysAspLeuThnGluGlnGluLysLIESerLys					872
Db	3172	ACACTTAATCGAGATGTTAAAGATTAATGAGATTAATGAGCAACAAAGCAAGATATTTTCT					3231
QY	873	---ProAspMetSerArgLeuArgLeuLaIleGlySerLaIleValLysLeuLa					890
Db	3232	TTAATACAGGAAAAATGAATCTCAACAAATGTTACAGACTGTTATATGCAAGAAAGAA					3291
QY	891	Gln-----GluProCysTyrHisGluLleLleThnLeuGlnTyr					904
Db	3292	CAATTGAAGACTGACTAAAGCAAAATATTCAATTGACACTTGAAMACAGAGAAATTA					3351
QY	905	GlnLeuCysLaIleLaIleAsnAspGluCysTyrGlnValArgGlnValPheLaGln					924
Db	3352	AGACTT-----CTTGGGGATGAACTTAAAAACCAACAAAGAGATAGTTGCACAA					3399
QY	925	LysLeuHisLysGluLeuSerArgLeuArgLeuProLeuGlnTyrMetLaIleCysLa					944
Db	3400	GAAAGAGACCTGCCATTAAAGAAA-----GAAGAGACCTTTTAGACCTGTGC					3450
QY	945	LeuCysLa-----LysAspProValLysGlnArgLaIleHisLaIleArgGlnCysLeu					962
Db	3451	AGACTGGCAGAGTGAAGAAAAAACTTAAGAAAAAGAGCCAGCAACTCCAAAGAAAAACAG					3510
QY	963	ValLysAsnLIEAsnValArgArgGluTyrLIEuLysGlnHisLaIleValSerGlu---					981
Db	3511	CAACAACCTTTAATGTATCAAGAAAGAGATGATGATGCAAGAAAAAGATTAATGAATA					3570
QY	982	---LysLeuLeuSerLeuLeuProGluTyrValValProTyr					994
Db	3571	GAGAAATTAAGATGAATTAAGACAAAGAAATGACATTGGAAACATATAGAAACGAG					3630
QY	995	ThrLIEHisLeuLeuLaIleHis---AspProAspTyrValLysValGlnAspIle---Glu					1012

```

Db 3631 AGCGTTGAGTTGGCTCAGAAACCTTAATGAAATTTATGAGAGTGAATCTATTAACCAA 3690
QY 1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleuMetAlaLysAsn 1032
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3691 GAAAGAAAGTCTCAAAAGAA-----TTACAGAAGTCAATTGAAACAGAG 3735
QY 1033 GluAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3736 AGAGACCACCTTAGAGATATATTAAGAAATTAAGACCTACAGCCTCAACCAACGAAGA 3795
QY 1053 AlaglnLysProAspAlaLysMetAsnGluLysLeuTrpThrValLysAspValAla 1072
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3796 GAACATAAAATTTGCTCATATTCACCTTAAGAAACCAAGAAACTATTTGATGAACCTAGA 3855
QY 1073 MetAsnIleLeuMetSerLysSerThrThrLysSerLeuGluSerProLysAspProVal 1092
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3856 AGAAGCGTA-----TCTGAGAAGACA----- 3876
QY 1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrpLeu 1112
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3877 -----GCTCAATTAATAATTAATCTACAGACTTAGAAAAATCCCATACCAA-----TTA 3924
QY 1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3925 CAAGAAAGATCCCA----- 3948
QY 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer-----Ser 1150
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 3949 GAGGAACAAGATTAATGCTCAATGTAAGAAAAAAGTCAAGTCAAGTCAAGAAACAATGAT 4008
QY 1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4009 GAACGTGAGTTATTAACAGAACAGTCCACAAACCAAGCTCAACAACTGAGGCAAGATA 4068
QY 1171 LysGlyArgLeuAspSerSerLysMetAspHisSerGlu-----AsnGluAspTrpThrMet 1189
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4069 -----GAAATGAAAGGCTCAGCGTTGATGTAAGTAATAATTTCAAGAA 4107
QY 1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerSerAspLeuValArgSer 1209
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4108 AGTCAGAGAGAGATTAATCTCTTAACCAAGAAAGAACAAACCTTAAGATTAAGAA 4167
QY 1210 GluLeuGluLysProAlaGlyArgLysLysThrProValThrGlu----- 1224
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4168 GCCCTTGAAATTAAACATGACACGCTGAAGAACATTTAGAGAACTTGGCTAAATC 4227
QY 1225 -----GlnGluGluLysLeuGlyMet-----AspAspLeu 1234
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4228 CAGGAGTCTCAAAAGCAACAAAGACAGCTCTTAATATGAAAGAAAAAGACAAATGAAC 4287
QY 1235 ThrLysLeuValGlnGluGln-----LysProLys----- 1244
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4288 ACCAAATCGTAGTGAAGTGAAGCATTCAAACCCAAAGTTCAGACACTAAGATA 4347
QY 1245 -----GlySerGlnArgSerArgLysArgLysHis----- 1254
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4348 GAATATGAAATGCTCGATTGTCCAAAAGACCTTCAGAAAGTCAATGAAATGAAATCT 4407
QY 1255 ThrAlaSerGluSerAspGluGln-----TrpProGluGluLys 1268
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4408 GTAGCTTAAGAGAAAGATGACCTACAGAGCGCTGCAGACAGTCTTCAATCTGAAGTAC 4467
QY 1269 ArgLeuLysGluAspIle-----LeuGluAsnGluAsp----- 1279
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4468 CAGCTCAAAAGAAACATTAAGAAATTTGATGCTTAACACCTGGAAACTGAGAGGAACCTT 4527
QY 1280 -----GluGluAsnSerProProLysLysGlyLysArgGly 1291
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4528 AAAGTTGCTCATTTGCTGCTGAAGAGACAGAGAACTATTAATAGAGTGAAGTGAAT 4587
QY 1292 ArgProProLysProLeuGlyGlyGlyThrProLysLysGluLysProThrMetLysThrSer 1311
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4588 CTTTCAGAGAAG-----GAAACTGAATATATCAACCATTCACAAAGCAG 4629

```

```

QY 1312 LysLysGlySerLysLysSerGlyProProAlaProGluGluGlnGluGlu 1331
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4630 TTAGAACATTCATGATGATTAATTAACAGAACAGATCCAAAGATTTATGAAAGAGAA 4689
QY 1332 -----ArgGlnSerGlyAsnThrGluGlnLys-----SerLysSerLysGlnHis 1346
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4690 CAACCTTAATTAATAACAAATTAAGAGCTTCAGGAAATAAGTGAACCTGAACCAATTTC 4749
QY 1347 ArgValSerArgArgAlaGlnGlnArgAla-----GluSerProGluSerSerAlaIleGlu 1365
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4750 AAGGAGCATGCAAAAGCCAAAGATTCAGCACTCAAGATATAGAAAGTAAAGATGCTGAG 4809
QY 1366 SerThrGlnSerThrProGlnLysGlyArg----- 1375
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4810 TT-GACCAACAGACTTCACAAAGATCAAGAAAGAAATTAATATGATTAAGAAAAAGA 4868
QY 1376 -----GlyArgProSerLysThrProSerProSerGlnProLys 1388
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 4869 GAAATGAAAGAGTACAGAGGCCCTTCAGATAGAGAGACCAACTGAAGAAAAA 4925

RESULT 3
US-09-595-684B-30
; Sequence 30, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Valsberg, Eugene
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cyto0036
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/595,684B
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 8257
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-30

Alignment Scores:
Pred. No.: 9,88e-11 Length: 8257
Score: 229.00 Matches: 285
Percent Similarity: 35.61% Conservative: 268
Best local Similarity: 18.35% Mismatches: 545
Query Match: 3.18% Indels: 455
Gaps: 66

US-09-512-581B-2 (1-1391) x US-09-595-684B-30 (1-8257)
QY 17 ProGlyValLysGluIleSerAspLysIleSerLysGluLysLeuValArgArgLeuLys 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1087 CCTATGTTAATGAGGTATCA-----ACTGATTAACCTCTCGTGAAGAGTATAGA 1137
QY 37 MetValAlaLysThrPheMetAspMetAspGlnAspSerLysGlu----- 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1138 -----AAAGAAATTAATGATCTTAATAAAACAAATTAAGAGAGGTTCTTATAGAGACG 1188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 52 -----GluLysGluLeuTrpLysAsnLeuAlaLeuHisLeuAlaSerAsp 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1189 CGGCGTCAGCAATGCAAAAGACCAATGCGCCAACTTTGGAAGAA-----AAAGAT 1242
QY 67 PhePheLeuLysHisProGlyLysAspVal-----ArgLeuLeuValAlaCys 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1243 TTGCTTCAGAAAGTACAGAAATGAGAAATTTGAAAACTTAACACGAGATGCTGTG----- 1296

```

OY	83	CysLeuAlaIaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys	102
Db	1297	-----ACCTCTCTCC-----	1308
OY	103	LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuAlaSprThrLysSer	122
Db	1309	-----CTCAGCTTGCAG-----	1333
OY	123	ProGlnPheAsnArgTyrPheTyrLeuLeuGlnLeuAlaIleAlaTrpValLysSerTyrAsn	142
Db	1336	AAAGAAACAGACAGATTACTTGCGCTTGCAAAATTAAACAAATAGACACTCAAC	1399
OY	143	IleCysPheGluLeuGlnAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe	162
Db	1396	-----TATGACGATCAATT	1410
OY	163	SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer	182
Db	1411	AATATACCAACAAT--ATAACAACAAAAACACATTAAGCTTTCATTAATTAATATACGA	1467
OY	183	SerIle-----IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal	199
Db	1468	GAATTTGATGATCGTGTGTGTACAGCTGTGATGTTTTCAGTAACTTGTATCATTA	1522
OY	200	Leu--ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla	218
Db	1528	AGTGAGATAGATGGATGCCAGAACCAAGCTACTAATATCAGGACAAATATGAA-----	1583
OY	219	LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn	238
Db	1561	-----	1581
OY	239	GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeu	258
Db	1582	-----AGTCAGTTGAACTCAGCTTCTCGTGCAGCTATGATATATGTG	1620
OY	259	IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu	278
Db	1621	GTATTA-----GACTATGACAACTGACCAAGAAAGAAAGAAATGGA	1665
OY	279	PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys	298
Db	1666	TTGGAATTTAAAGAAAGAAATGATTTGGATGAATTTGGCTCTGAAAGAAACCTAAA	1725
OY	299	MetPheGlyAlaLysAspSerGlnLeuLeuLysSerGlnAsnLysProLeuThrGlnCysTyr	318
Db	1726	-----AAGATCAAGAGATGCA-----	1743
OY	319	LeuGluArgPheAsnAspIleHisValProIleArgLeuGlnCysValLysPheAlaSer	338
Db	1743	-----	1743
OY	339	HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg	358
Db	1744	-----CTAATTCATGAAATTTCCAGACTTAAGAATTAAGTTAAGCACTGCAAGATATAT	1797
OY	359	SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla	378
Db	1798	AATCAAGATCTTGGAGATGAATC-----AGTTCA	1827
OY	379	LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu	398
Db	1828	AAAGTAGAGCTGCTTAGAGAAAGACAGACCAAGATTAAGAAAGATACAGATACATGAC	1887
OY	399	AspLysArgTrpArgValArgLysGluAlaMetCketGlyLeuAlaGlnIleTyrLysLys	418
Db	1888	TCTCAAAAG-----CTGAAAAATTAATAATGCACTTGCTGA-----	1923
OY	419	TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp	438
Db	1924	TATCATTTGGAAAGCATTAAGAACCCAAACAAATGAATGACAGACTCTGTGTGATGCTGA	1983
OY	439	LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle	458

[illegible]

Db 2920 ---GAGAGGACCAACTCAAAAGTGAATTCACGATACCTTTAACAATGAATAGACT 2976
 QY SerAsn-----LeuGlnHisLeuIleThrProLeuValThrIleGlyHis 772
 Db 2977 CAGAACAAATTAAGCAATGCTTGAAGTCTGAAACACATCAAGAAACAAATTAATACA 3036
 QY IleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLysSerThrPheAlaThrPhe 792
 Db 3037 CTA-----AAATCGAAAAATTTCTGAGAA 3060
 QY IleValLysAspLeuLeuMetLysAspArgLeuProGlyLysThrThrLysLeuTrp 812
 Db 3061 GTTTCACGAAATTTGCATG----- 3081
 QY ValProAspGlnGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
 Db 3082 -----GAGGAAATTAACAGGAAACCTAAAGATCAATTTCTG----- 3117
 QY ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
 Db 3118 ---CAAAAGATGGTTGGCATAGATMAAAACAGAGATTGGAAAGCTAAAAATATCC---CAA 3171
 QY LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLys 872
 Db 3172 ACACCTAACTGCAGATGTTAAGATTAATGACATAATTGACACAAAGAGATATTTTCT 3231
 QY 873 -----ProAspMetSerArgLeuArgLeuAlaIleAlaIleValLysLeuAla 890
 Db 3232 TTATATACAGAGAAAAATGAATCCCAACAATGTTAGAGTGTTATAGCAGAAAGAA 3291
 QY 891 Gln-----GluProCysTyThrHisGluIleIleThrLeuGlnGlnTyr 904
 Db 3292 CAATTGAAAGACTGACCTMAAGAAAAATTTGMAATGACCATTTGMAAACCGAGAAATTA 3351
 QY 905 GlnLeuCysAlaLeuAlaIleAsnAspGlnCysTyThrGlnValArgGlnValPheAlaGln 924
 Db 3352 AGACTT-----CTTGGGATGAACCTTAATAAGACAGAGATGTTGCACAA 3399
 QY 925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTyrMetAlaIleCysAla 944
 Db 3400 GAAAGAACCATGCGCATTAAGAA-----GAAGAGAGCTTTCTAGAGACTGTGAC 3450
 QY 945 LeuCysAla-----LysAspProValLysGlnArgAlaHisAlaArgGlnCysLeu 962
 Db 3451 AGACTGCGAGAGTTGAAGAAAACTTAAGAAAAAGCCAGCAACTCCAAAGAAAAACAG 3510
 QY 963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleValSerGlu--- 981
 Db 3511 CAACAACCTCTTAATGTACAGAAGAGATGATGATGCCAGAAAAAGATTATGAATA 3570
 QY 982 -----LysLeuLeuSerLeuLeuProGluTyrValProTyr 994
 Db 3571 GAGAAATTTAAAGATGAATTAAGAACAAAGAAATGACATTCAGCAATGAGAAACAG 3630
 QY 995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
 Db 3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAATAATTAATGAGAAAGTATATACCAA 3690
 QY 1013 GlnLeuLysAspValLysGlnCysLeuTrpPheValLeuGlnIleLeuMetAlaLysAsn 1032
 Db 3691 GAAAGAAAAAGTCTAAAGAA-----TTACAGAAAGTCAATTTGAAACAGAG 3735
 QY 1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAsp 1052
 Db 3736 AGAGACACACCTTAGAGATATATTAAGAAATTAAGCTACAGCCTCAACAAACAAAGAA 3795
 QY 1053 AlaGlnGlyProAspAlaLysMetCAsnGlnLysLeuTyrThrValCysAspValAla 1072
 Db 3796 GAACTAAAAATTCCTCATATTACCTAAAGAACCCAGAAACACTTATGATGAACCTAGA 3855
 QY 1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlnSerProLysAspProVal 1092
 Db 3856 AGAAGCGTA-----TCTGAGAGACACA----- 3876

QY 1093 LeuProAlaArgPheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
 Db 3877 -----GCTCAAAATTAATTAATCTCAGAGACTTGAAAAATCCCATACCAAA-----TTA 3924
 QY 1113 ProProGlnMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
 Db 3925 CAGAAAGAGATCCCA-----GTGCTTCAT 3948
 QY 1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer---Ser 1150
 Db 3949 GAGACACAGAGATTACTGCTTAATGTGAAAAAGTCAAGTCAAGCTCAGCAACAAATGAAT 4008
 QY 1151 ArgMetGlnThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
 Db 4009 GAACGTGAGTTATTAACAGAACAGTCCACAAACAGACTCAACAACTGCAGCAAGATTA 4068
 QY 1171 LysGlyArgLeuAspSerSerGlnMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
 Db 4069 -----GAAATGGAAGAGCTCAGCTTGATGAATAAAATTTCAAGAA 4107
 QY 1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
 Db 4108 AGTCAGACAGATAAATATCTTAACCAAGAAAGACACACTTAACAGATAAAGAA 4167
 QY 1210 GlnLeuGlnLysProArgGlyArgLysLysThrProValThrGlu----- 1224
 Db 4168 GCCCTGAGATTAAACATGACCAAGCTGAAAGAACATATTAGAAACCTTGCTAAATAC 4227
 QY 1225 -----GlnGlnGlnLysLeuGlyMet-----AspAspLeu 1234
 Db 4228 CAGAGCTCTCAAGCAAAACAAACAGTCCCTTAATTAATGAAGAAAAAGCAATGAACCT 4287
 QY 1235 ThrLysLeuValGlnGlnGlu-----LysProLys----- 1244
 Db 4288 ACCAAATTCGTGAGCTGAGTGAAGCAATTAACCCCAAAATTCAGACACTTAAGATA 4347
 QY 1245 -----GlySerGlnArgSerArgLysArgLysHis----- 1254
 Db 4348 GAATATGAATATGCTCGATGTTGCCAAAGACCTTCAAGAAAGTCATATGAATGAATCT 4407
 QY 1255 ThrAlaSerGlnSerAspGlnGln-----TrpProGlnGlnLys 1268
 Db 4408 GTAGCTAGAGAAAGATGACACAGAGCTGCACAAAGCTTTCATCTCAAGTGAAC 4467
 QY 1269 ArgLeuLysGlnAspIle-----LeuGlnAsnGlnAspGlnGln 1281
 Db 4468 CAGCTCAAAAGAAACATTAAGAAATTTAGCTTAACACCTCGAACTGAAGAGAACTT 4527
 QY 1282 AsnSerProProLysLysGlyLysArgGlyArgProPro---LysProLeuGlnGlyGly 1300
 Db 4528 AAAGTTGCTATTGTTCTCTGAAAGAAACAGAGAAACCTATTAAGTTAAGTGAAT 4587
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLys-----GlySer 1315
 Db 4588 CTTTCAGAGAGAAAGAACTGAATTAACCACTTCAAAAAGGAGTTAAGACAAATCAATGAT 4647
 QY 1316 LysLysLysSerGlyProProAlaProGlnGlnGlnGlnGlu-----GluArg 1332
 Db 4648 AAATTTACAGAAACAAAGATCCAGAGATTTATGAGAAAGAGAACAACTATTAATAACAA 4707
 QY 1333 GlnSerGlyAsnThrGlnGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
 Db 4708 ATTAGTAGGTTGAGGAAAGAGAACTGAATGAACCAATTAAGAGAGATCGCAAAAGCC 4767
 QY 1353 GlnGlnArgAla---GluSerProGlnSerSerAlaIleLeuSerThrGlnSerThrPro 1371
 Db 4768 AAGGATTCACACTACTCAAAAGTATAGAAAGTATGATGCTGAGTT-GACCAAGAGCTTCA 4826
 QY 1372 GlnLysGlyArg----- 1375
 Db 4827 AGAAAGTCAGAAAGAAATTAACAATTAATGATTAGGAAAAAGAGAAATGAAGAGACTACA 4886

QY 1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
||| ||||| |||
Db 4887 GGAGGCCCTTCAGATAGAGACAGACCACTGAAGAAAA 4925

RESULT 4
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4

Alignment Scores:
Pred. NO.: 9.22e-09 Length: 6755
Score: 207.00 Matches: 78
Percent Similarity: 42.81% Conservative: 50
Best Local Similarity: 26.09% Mismatches: 137
Query Match: 2.88% Indels: 36
Gaps: 9

US-09-512-581b-2 (1-1391) x US-08-931-999-4 (1-6755)

QY 1121 ThrProGlyLysProLys---ThrThrsnValLeuGlyAlaValAsnLysProLeuSer 1139
||| ||||| ||||| |||||
Db 5664 ACAGCACCAAAACCCAGACAGACGACAAAGAAAAAGACAAACCAAGACAA 5723

QY 1140 SerIaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSer 1159
||||| ||||| ||||| |||||
Db 5724 GAAAAAGAAAAACAAAGACAAAGAAAAAGAAAAAGAAAAAACACGACGCCGA 5783

QY 1160 SerSerSerAsnProSerSerProGlyArgGlyLysGlyArgLeuAspSerSerGlnMet 1179
|||::: ::|||::: |||

Db 5784 CAGAAAGACAGACGAGGAAACGACACACAGGACACAGGACACACCCAGAAACA 5843
QY 1180 AsphHisSerGluAsnGluAspTyrThrMetSerPro---LeuProGlyLysLysSer 1198
::: |||||
Db 5844 CCAACACACAGAACCAAGAGAAAGACAGACGACAAACCAACACAGGAGAGGAAA 5903

QY 1199 AspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgLys 1218
||| ||||| ||||| |||||
Db 5904 ACCAAAGAGAGAACCAAAAG-AAAAAGCCACGACGACCAAGAAACCAAGAAAGGAGAA 5962

QY 1219 Lys-ThrProValThrGlnGlnGluGlu-----LysLeuGlyMetAspAspLeuTh 1235
||| ||| ||||| ||||| |||||
Db 5963 AACACGAAACGAAAAACAAAGAGAAAGAAAGACAAACAAAGAGAGACGAGAAACA 6022

QY 1235 LysLeuValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisTh 1255
::: ||||| ||||| ||||| |||||
Db 6023 AAGAAACAGAGAACGACAGAAAGAAAGACAAACAAACAAAGAGAGGAGAGAGAA 6082

QY 1255 LAlaSerGluSerAspGluGlnGlnThrProGluGluLysArgLeuLysGluAspIleLe 1275
||| ||||| ||||| ||||| |||||
Db 6083 GGAGCAGGAAAA-GACCAAGAAAAACACACACGAAACGAAAGCAGCGAGA----- 6130

QY 1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLy 1295
||| ||||| ||||| ||||| |||||
Db 6131 -GAAGCAGACGAAGAAAGAAAGAAAGACCAACCAACCAACCAACGAGACAGACAAA 6189

QY 1295 sProLeuGlyGlyGlyThrProLysGlnGlnGluProThrMetLysThrSerLysGlySe 1315
| ||||| ||||| ||||| |||||
Db 6190 GACCAAAACAAACACGAAAAACGACAGAAACAAACACAAAGAGAGAAAAAGAAAAGA 6249

QY 1315 rLysLys-----LysSerGlyProPr 1322
||||| ||||| ||||| ||||| |||||
Db 6250 AAAAAACGACGAGAAACAAAGACGAGGAAACAAACAAACCCGCAAGACGAGCC 6309

QY 1322 oAlaProGluGlnGluGlnGluGlnArgGlnSerLysAsnThrGlnLys----- 1340
||| ||||| ||||| ||||| ||||| |||||
Db 6310 CACAAAGAGAGAAAGACCGGAAACCAACCAACCAACCAACCAACCAACCAAGAAC 6369

QY 1341 ---SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerPr 1359
||| ||||| ||||| ||||| ||||| |||||
Db 6370 AACAAAGCAAAAGAAAGAAAGAGGAAAGAAAGAAAGAGAAACGAGACAAACAAACCC 6429

QY 1359 o-----GluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGly---ArgG 1376
||| ||||| ||||| ||||| ||||| |||||
Db 6430 AAAAAACACAGACGAGGAGAGAAAGACCAACCAACCAACCAACGCGCAAAAGGCCAAAAAGCC 6489

QY 1376 YArgProSerLys-----ThrProSerProSerGlnProLysLys 1389
||||| ||||| ||||| ||||| ||||| |||||
Db 6490 ACAGCAACAAACAAAGACAAACAGACACAAACAAAGGAGAAACACACACCCAAAAAAA 6544

RESULT 5
US-08-353-700-2
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSLIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORPMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,700
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: REED, JANET E.
 REGISTRATION NUMBER: 36,252
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10136 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: HUMAN
 US-08-353-700-2

Alignment Scores:
 Pred. No.: 6,44e-08 Length: 10136
 Score: 201.50 Matches: 283
 Percent Similarity: 35.16% Conservative: 242
 Best Local Similarity: 18.96% Mismatches: 587
 Query Match: 2.80% Indels: 382
 DB: 1 Gaps: 63

US-09-512-581b-2 (1-1391) x US-08-353-700-2 (1-10136)

QY 19 Vallysgluileseraspysile-----serlysglu 30
 Db 5868 ATGAAGAATTACACTCAAAACCTCCATTACAGAGGTACACTAATGACCAAAATTGAA 5927
 QY 31 Metvalargyleuysmetvallysthrphenetaspmetaspglaspserglu 50
 Db 5928 GCATGCATACAGATTGGAAAAAATGTTGGGAACTTAAGAAAGAAACCTCAGATTTAAGT 5987
 QY 51 GlugluysgluleuTYR----- 56
 Db 5988 GAAAAATGGAAATATTTTCTTGATCCACCAGAGTTACTCCAGAGTAGAAGACTTCT 6047
 QY 57 -----LeuasnleualaleuHISleualaseraspheleuysHISproglylys 74
 Db 6048 GAAGGCGCTCAATTCGTGATTGAGAAATGCATGCAGAT-----AAATCATCAGCTGAA 6098
 QY 75 AspvalargyleuValalaCysCysleualaspheargileTYRAlaProglu 94
 Db 6099 GATATTGGAGATATGAGCC----- 6119
 QY 95 AlaProTYRThrserProaspysileuysaspillephemethelethrargglu 114
 Db 6120 -----AAGTGAATGCACAGCTGGAAGAGAGATTCTTGATGTGGAAAAATGAGCTG 6170
 QY 115 llysglyleuyluspthrlysserProgluInpheasnargTYRphenyTYRleuylusasn 134
 Db 6171 AGTAGATTCAGATCGGAGAAAGCTAGCATTCAGCATGAAGCCCTACCTCGAGAGCTGAC 6230
 QY 135 llaalatrpVallysserTYRasnIleCysPhegluleuyluspsersasn-----Glu 152
 Db 6231 TTGAGAGTAGTCAACACAGAGAGCTATGTTTAGAAAAAGCAATGAAATTAAGCAGAG 6290
 QY 153 llaPhehthrGlnleuTYRargTYRThrPheSerValIleasnnglyHISasn----- 170
 Db 6291 GTTATTGTCTGCCCTTGAAGAAGACTC---TCAGTGTCAAGTGAAGAAACACAGCTT 6347
 QY 170 ----- 170
 Db 6348 CGTGGAGATTAGATCTATGTCAAAAAAACACAGCGCACTGGATCAGTTGTCTGAAGAA 6407

QY 171 -----GlnlyValHISmethISmetValaspmetSerIle--- 184
 Db 6408 ATGAAGAGAAAAACACAGAGAGCTTGACTCTCAAAAGTGTCCATTCATTCAG 6467
 QY 185 llecysgluclyaspthrValserGln-----GlnleuysaspthrValleuValasnleu 203
 Db 6468 GTGGCAGAGCAGAGGAGGAAAGAAAGAGCACTCTTACAGCTTGTCTCTGATGTG 6527
 QY 204 ValProIaHISlysaSnleuysGlnAlaTYRaspheulalysalaleu----- 221
 Db 6528 -----AGTAGCTGTAAAGAAACAACTCATCTCCAGAGAAAGCTGCAGAGT 6575
 QY 222 LeuysargThrAlaGlnAlaIleGluProTYRThrPhePheasnGlnValleu 241
 Db 6576 TTGGAAAAGAGCTCAGAGCAGCTCTTTCACAAATGTGAGCTGGAAGAAACAAATGCA 6635
 QY 242 Metleuyllys-----ThSerIleaseraspseuSerGlnHISVal 255
 Db 6636 CAACCTGAATTAAGAGAAAGATTGCTTGACAGGAATCTGAAAGCTGCAGGCCAGACTG 6695
 QY 256 PheaspheulleuylusleuTYRAsnIleaspserHISleuylusleuSerValleuPro 275
 Db 6696 AGTGAATCAGATTATGAAAGCTGAATGTCTCCAGAGCTTGAGAGCGCAGCTGTGAGAG 6755
 QY 276 GlnleuylusleuysleuysSeraspaspasnleu-----Argleu 290
 Db 6756 AAAGGTGAGTTCGATGAGAGCTGAGCTCAACACAGAGAGAGTGCATCAGTAGAGAGA 6815
 QY 291 GlnValAllylsleuValAlaCysMetPheGly-AlaIlyaspserGlnleuylasergl 310
 Db 6816 GGCATCGAGAACTGAGAGTTCGATTGAGCGCCGATGAAAGAAAGAGCTGCACATTCGA 6875
 QY 310 nasnlyspProleuTYRglnCysTYRleuylusArgPheasnAspIleHISValProIlearg 330
 Db 6876 GAGAAAC-----TGAAGAAGCGGAGCGGAGATG-----ATTCACTTAAGATGATAA 6923
 QY 330 gleuylusVallysphealaserHISlys-----le 341
 Db 6924 GTTGAGAACCTTGAAAGGAATTCAGATCTGAGAAAGAAACAGAGAGTATGATTCTT 6983
 QY 341 uMetAsnHISProaspheulalysaspheulThrGlnTYRleuysValargserHISas 361
 Db 6984 GATG-----CCGAGATTCCAAGGCAGAA-GRAGAACTCTAAAACACAAATAGAGAGA 7036
 QY 361 ProgluGlnAlaIleargHISaspyAlaIleValserIleValThrAlaAlaIlysyas 381
 Db 7037 GATGGCCAGAGAGCTGAAAGATTTTGATTAAGACCTGTGACGTTAAGCTGAAAGAGA 7096
 QY 381 pIleuylusValasn-----aspHISleu 390
 Db 7097 AAATCTGACAAAACAAATACAGAAAAACAGAGTCTGCAGAACTAGACAAAGTTACT 7156
 QY 390 uasnPheValargGluargThrleuaspysileargTYR-----ArgValarglysgl 407
 Db 7157 CTCTTCATTAAAAAGTGTGTAGAGAAAAAGGACAGAGATACAGATCAAGAGAGA 7216
 QY 407 uAlametleuGlyleuAlaGlnIleTYRlys-----LysTYRAlaGlnleuInsAlaI 425
 Db 7217 ATCTAAAACTGAGTGAAGATGCTTCAGATCAGTTAAGAGAGCTAATAGAGCAGTAGC 7276
 QY 425 aglylysaspAlaAlaIlysglnIlealatrPlelyaspysleuylusleuHISleuTY 445
 Db 7277 AGCCTGTGTGTGGCCAAAAATTAAGAGGCCAGCAAGACAGTGTACACCCACCAAT 7336
 QY 445 rglnAsnserIleaspasprargleuValGluargIlePheAlaGlnTYRmetValpr 465
 Db 7337 AGAGGAAGACATCAGCTGGAATATGACATTGAAAGCTGAGAGCCGCG----- 7385
 QY 465 oHISasnleuGlnThrThrGluargmetlyscysleuTYRtyrleuTYRAlaThrleuas 485
 Db 7386 -----CTAGAGGCTGATGAAAAAGAGCAG----- 7409
 QY 485 pleuasnAlaVallysalaleuasnGlnumetTYRlyscysGlnasnleuValargHISgl 505

Db	7410	-CTTGTGCTCTTACACACTGAAAGAAAGTGCACATCATCAATTTACTTAAGGGTAG	7468
Oy	505	nvallylsAspLeu-----IleAspLeuIleLysGlnProLysThrAspAlaSerVa	522
Db	7469	AGTGGAGAACCTTGAAGAGAGCTTACAGATGACGAGCAAAACAGAGCATGCAGCTCT	7528
Oy	522	llyAlaAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl	542
Db	7529	TGAGGCGAGAGATTTCCAAAGCAGAGGTGAGAGACCTAAAGCAAAATATGACGAGATGAC	7588
Oy	542	aglnAspPhe-----MetLysLysPhe	549
Db	7589	CCAAAGTCCTGAGAGCTCGAATTAGATGTTTACTATTAAGTCAGAAAAAGAAATCT	7648
Oy	549	eThrGlnValLeuLysAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr	569
Db	7649	GACAAATGAATTTACAAAAGAGCAAGACCGAATATCTGAATTTAGAAATATTAATTTCA--	7706
Oy	569	oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs	589
Db	7707	-----TCATTTTGAAATATTTTCTTCAAGAAAAAGACAGAAAGATGACA	7750
Oy	589	nProLysGlnProThrAsnProPheLeuGlnMet-----IleLysPheLeuLe	605
Db	7751	GATGAAAGAAAAATCAACAGCACTGCCATGAGAGATCTTCAACACAAATTTAAAGAGCTCAA	7810
Oy	605	uglArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLysGlnVa	625
Db	7811	TGAGAGAGTGGCACCCTCGCTATATGACCAAGAAAGCCCTGTAAGGCC-----AAAGACA	7864
Oy	625	IAsnLysSerIleAspGlyThrAlaAspAspGluAspGluLysAlProThrAspGlnAl	645
Db	7865	GAATCTTAGT-----AGTCAAGTAGAGGTGCTTGAACTTGAAGAGGC	7906
Oy	645	a-----IleArgAlaLeuGluLeuGluValLeuSerPheThrHisProIleSerPh	664
Db	7907	TCAGTTCTCTCAAGACCTTGATGAGGCCAA-----AATAATTATATTTGTTT	7954
Oy	664	eHisSerAla-----GluThrPheGlnSerLeuLeuAlaLysLeuLysMe	679
Db	7955	GCATCTTCACTGAAGAGCCCTCATTTCAAGAAGTAGAAGATGCGCAAGCAAACTGGAGAA	8014
Oy	679	tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysII	699
Db	8015	GAAGGATGAGAAATTCAGTACATGAAATAATCAAT-----	8051
Oy	699	eGluGlnAspPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis--	716
Db	8052	-----CAAGACCAGAGCAGCTTGTCTCTAAATCTCCAGGTGGAGAGCAGCACCACT	8107
Oy	717	-----LysSerLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl	735
Db	8108	TTGGAGAGAGCAAACTTAGAAGTACAGAAATCTGACGGTGAATTTGAGCAGAAAGATCA	8167
Oy	735	alIlePheSerSerLysGlnThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe	755
Db	8168	AGTCTTCAATCCAAATAAGCGCTTTTGCGAGACACATTTAGAAATGCTCGCAATTTCTTA	8227
Oy	755	uAspProSerAsnLeuHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe	775
Db	8228	C-----AAGATCTTAGAGAAAT-----GAGCTTGAATT	8254
Oy	775	uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy	795
Db	8255	GACAAATAATGCAAAATG-----TCCTTTTGT-----	8282
Oy	795	sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs	815
Db	8283	-----GAAAAAGTAAACAAATATGACTGCAAAAGCA	8311
Oy	815	pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr	835

Db	8312	AACTGAGCTCCAGAGGCAAAATGCAATGACATGGCCACAGAAAACAGCAGACTGCAGAAAGA	8371
OY	835	pleuenclymetylsasn-----asnHisSerLy	845
Db	8372	ACTCAGTGGCAGAGAAAATTAGCTAGCTGGAGAGAGTTGCAGTTACTGTTGCAAGAAATATA	8431
OY	845	sSerGlyThrSerThrLeuArgLeuLeuThr-----	855
Db	8432	GAGCAGCAAAATCATATTGGAAGAGAGCTCACACTAGAAAATAGTGAATTGGAAGAGAGCCT	8491
OY	856	-ThrLeuLeuHisSeraspGlyaspLeuThrGluGlnGlyHisSerLysProaspMe	875
Db	8492	AGATTGCATGCACAAAGAC---CAGCTGAAAAAGGAAAGGAAGTG---AAGAGAGAAAT	8545
OY	875	tSerArgLeuArgLeuAlaIaGlySerAlaIleValLysLeuAlaGlnGluProGlyCys	895
Db	8546	AGCTGAATATTCAGCAGCTACGGCTTCATGAAGCTGAAGAAAGAA-----	8588
OY	895	rHisGluIleIleThrLeuGluGlnGlyTrpGluLeuCysAlaLeuAlaIleasnaspGlyCys	915
Db	8586	-CACCAGCGCTTTGGCTTTTGGAC-----ACAAACAAACA	8612
OY	915	sTyrglnValArg---GluValPheAlaGlnLysLeu-----HisLysGlyLeu	931
Db	8618	GTATCAATCAAAATCCACAGACATCCGAGAGAAATGTGACTCTTAAGAAGAAATGTCTCAG	8677
OY	931	rArgLeuArgLeuProLeuGluGlyTrpMetAlaIleCys-----AlaLe	945
Db	8678	TTCCACAGAGAGCTGGAGATAGACCTTTTAAAGCTGATTAAGACAGCTCAATTAATTCATT	8737
OY	945	uCysAlaLysaspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs	965
Db	8738	GAAAGCTACTACTCAGATTGTTGGAAGATTGAAGAAACCAAGATGAGCAATCTAAATA	8797
OY	965	nIleasnValArgArgGluGlyrLeuLysGlnHisAlaIleValSerGluLysLeuLeuSe	985
Db	8798	TGTAAAT-----CAGTTTAAACAGAAATATGACCTGGCCCGAGGGAATATPAA	8845
OY	985	rLeuLeuProGluTrpValValProGlyThrIleHisLeuLeuAlaHisaspProaspTy	1005
Db	8846	GTTGTGTC-----	8852
OY	1005	rValLysValGlnaspIleGluGlnLeuLysaspValLysGlnCysLeuTrpPheValLe	1025
Db	8853	-----ATCAAAATCCTGTAACACAGCTGGAAGAGAAAGAGATACGCAGAAAGAACT	8905
OY	1025	uGluIleLeuMetAlaLysasnGluValasnSerHisAlaPheIleArg-----	1041
Db	8906	CTCTCAACTTCACAGCTGCACAGAGAAACAGAAACACAGTACTGTATGATACCAAGGT	8965
OY	1042	---LysMetValGluValnIleLysGlnThrLysaspIaGlnGlyProaspaspAlaLy	1066
Db	8966	CGATGAATTAACAACCTGAGATCAAAAGACTGAAGAATACTCTTGACAGAAAAACCAAGCA	9025
OY	1066	sMetAsnLysLeuLysrTrpThrValCysaspValAlaMetAsnIleIleMetSerLysSe	1080
Db	9026	GGCAGATCAATACACTTGGATGACTGT-----TCCTTGCTTAATAAGCCATGA	9073
OY	1080	rThrThrTyriSerLeuGluSerProLysasp-----ProValLeuProAlaArgPhePh	1098
Db	9074	A-----AAGTTAGAGAAAGCTAAAGAGATGTTAGACACACAAAGTGGCCCATCTGTG	9124
OY	1098	eThrGlnProaspLysasnPheSerAsnThrLysAsnTrpLeuProGluGluMetLysSe	1118
Db	9125	TTTCACAGCAATCTAAACAAGATTCGCCGAGGCTTCCTTTGGTAGGCTCCAGTTGTT----	9179
OY	1118	rPhePheThrProGlyLysProLysTrpThrAspValLeuGluLysAlaValAsnLysProLe	1138
Db	9180	-----CCAGGACCACTTCACCAATCCCTTCGTGTACTGAAAAAGAGGTT	9220
OY	1138	uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl	1158
Db	9221	ATCATTTCTGGCAAAATATAAGCTTCACAGCAAGGCAAGGCAATCCAGTGAATATTTGGAGAA	9280

```

QY 1158 aSerSerSerSerSerProSerProGlyArgIleLysGlyArgLeuAspSerSerG1 1178
Db 9281 TGTAGAGAGACCAACACTGCTACCCA----- 9308
QY 1178 uMetAspHisSerSerLysAspGlyThrMetSerSerProLeuProGlyLysLysSe 1198
Db 9309 -----GAGACCTTTTCT----- 9328
QY 1198 rAspLysArgAspAspSerSerAspLeuValArgSerGluLeuLysProArgGlyArgLy 1218
Db 9329 CAAGAAAGCAGTCACTAGTGTATTCACCTGCAGAA----- 9365
QY 1218 sLysThrProValThrGluGluGluLysLeuGlyMetAspAspLeuThrLysLeuVa 1238
Db 9366 -GACACGGAAGGTACTAGTTGAGCAGAG--GCACTTCAGAGAGTTGAAGAAAG 9421
QY 1238 lGlnGluGlnLysProLysGly-----SerGlnArgSerArgLysArgLysIstHraI 1236
Db 9422 GTTTGCTGACATCCCGACAGAGAACTAGCCATATATCTCGCAGAAACACCATGGC 9481
QY 1256 aSerGluSerAspLysGluGlnIntPrProGluLysArgLeuLysGluAspLleLeu-- 1275
Db 9482 AACTCGACACGACCCCGCTGCTGCACAGAACTTAGCCTATCCCACTGAGTCTCG 9541
QY 1276 ----GluAsnGluAspLysGlnAsnSerProProLysLysGlyLysArgLysArgProPr 1294
Db 9542 CAAGAAAAATCTTCAGAGTCTCCAAACCAACAGCTGTGCGACAGATCACAAAGGT 9601
QY 1294 oLys-----ProLeuGlyGlyGlyThrProLysGluGluProThrMetLy 1309
Db 9602 CAAGTTGCTCAGCGGAGCCAGATGATTCAGGACCATCTCCGAGAACCCACACGAA 9661
QY 1309 s-----ThrSerLysLysGlySe 1315
Db 9662 ATCCGTCCAGTCAATATCTCTCTGAGAGAAAGTCCGACAGCCCAAGAGAGGCGCT 9721
QY 1315 rLysLysLysSerGly-----ProProAlaProGlu 1335
Db 9722 GAGGTCAGAGGAGCGGAGCTTTGTCACAGCCCAAG 9758

RESULT 6
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-2

Alignment Scores:
Pred. No.: 6,44e-08 Length: 10136
Score: 201.50 Matches: 283
Percent Similarity: 35.16% Mismatches: 242
Best Local Similarity: 18.96% Mismatches: 587
Query Match: 2.80% Indels: 382
DB: 5 Gaps: 63

US-09-512-581b-2 (1-1391) x PCT-US95-16216-2 (1-10136)

QY 19 ValLysGluLysSerAspLysIle-----SerLysGluGln 30
Db 5868 ATGAAGATATGAGTCACTCAAACTCCATTACAGGAGGTACAACTAATGACCAAAATTGAA 5927
QY 31 MetValArgArgLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
Db 5928 GCATGCATAGATTTGAGAAAAATAGTTGGGAACTTAAGAAAGAAACTCAGATTTAAGT 5967
QY 51 GluLysLysGluLeuTy----- 56
Db 5988 GAAAAATTTGAAATTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTGAATCTCT 6047
QY 57 -----LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
Db 6048 GAAGCCTCAATCTGATTGATTAAGAAATGATCAGAT-----AATCATCAGCTGAA 6098
QY 75 AspValArgLeuLeuValAlaCysCysLeuAlaAspLlePheArgIleTyAlaProGlu 94
Db 6099 GATATTGAGATATATGTGGC----- 6119
QY 95 AlaProTyThrSerProAspLysLeuLysAspLlePheMetPheIleThrArgGlnLeu 114
Db 6120 -----AAGTGAATGACAGCTGGAAGAGAGATTTCTTGTGGAAGAAATGAGACTG 6170
QY 115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyThrPheTyLeuLeuGln 134
Db 6171 AGTAGATCAGATCGGAGAAAGCTAGCATTCAGATGAAAGCCCTCTACCTGAGGCTGAC 6230
QY 135 IleAlaTrpValLysSerTyAsnIleCysPheGluLeuGluAspSerAsn-----Glu 152
Db 6231 TTAGAGTAGTTCAAACAGAGACGATATGTTTAAAGAAACAAATGAAATTAACACAGAG 6290
QY 153 IlePheThrGlnLeuTyArgTyThrLeuPheSerValIleAsnAsnGlyHisAsn----- 170
Db 6291 GTTATTTGTCCTTGAAGAGAACTC--TCAGTGTGACAAAGTAGAGAAACCACTG 6347
QY 170 ----- 170
Db 6348 CGTGAGATTAGATCTATGTCAAAAAACACAGCGCACTGATGAGTGTCTGAAAAA 6407
QY 171 -----GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
Db 6408 ATGAAGGAGAAACACAGAGCTTGTAGTCTATCAAAAGTGAGTCTTCATTCGATTCAG 6467
QY 185 IleCysGluGlyAspThrValSerGln-----GluLeuLeuAspThrValLeuValAsnLeu 203
Db 6468 GTGCGAGAGGAGAGGAGAGAAAGCAAGCAACTCTTCAGACTTTGTCTCTGAGGTG 6527
QY 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyArgAspLeuAlaLysAlaLeu----- 221

```

Db 6528 -----AGTAGCTGTATAAGACAAACATCTCCAGAAAAAGCTGAGACT 6575
 QY LeuysArgThrAlaGlnAlaIleGluProGlyrThrPhePheasnGlnValLeu 241
 Db 6576 TTGAAAAGAGACTCAGGACCTGCTTTGACAAAATGAGCTGAAAACCAATTTGCA 6635
 QY MetLeuGlyLys-----TherSileSerAspSerGlnHisVal 255
 Db 6636 CACGTGAATTAAGACAAAAGATTCCTTGCACAGAAATCTGAAGGCTGCAGGCCAGACTG 6695
 QY 256 PheAspLeuIleLeuGluLeuTyrrAsnIleAspSerHisLeuLeuSerValLeuPro 275
 Db 6696 AGTGAATCAGATTATTAAGAGCTGAATCTTCACAGGCTTGAGGCCGACATGGTGAG 6755
 QY 276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu-----ArgLeu 290
 Db 6756 AAAGGTGAGTTCGCTGAGGCTGAGCTCAACACAGAGGAGAGTGCATCAGCTGAGAGA 6815
 QY 291 GlnValValLysLeuValAlaLysMetPheGly-AlaLysAspSerGlnLeuAlaSerGln 310
 Db 6816 GGCATCGAAGAACTGAGATTCGCTTGAAGCCGATGAAGAGACAGCTGCACATCGCA 6875
 QY 310 AsnLysProLeuThrGlnCysTyrrLeuGlyArgPheAsnAspIleHisValProIleArg 330
 Db 6876 GAGAAAC-----TCAGAAACCGCAGCGGAGAAATG-----ATTCACTTAAGATTA 6923
 QY 330 GLeuGluCysValLysPheAlaSerHisCys-----Le 341
 Db 6924 GTTGGAACCTTGAAGGAAATTCAGATTCAGAAAGAACACAGAGCTAGTCTT 6983
 QY 341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrrLeuLysValArgSerHisAs 361
 Db 6984 GATG-----CCGAAATTCACAGACAGAA-CTAGAGACTTAAACCAAAATGAGAA 7036
 QY 361 ProGluGlnAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
 Db 7037 GATGCCAGAGCCCTGAGATTTTGTGAATTAACCTGTACAGTTAAGCTCGAAGAA 7096
 QY 381 PLeuLeuValAsn-----AspHisLeuLe 390
 Db 7097 AAATTCGACAAACAAATACAGAAACAAAGGTCAGTTCAGAACTAGACAACTTACT 7156
 QY 390 uAsnPheValArgGluArgThrLeuAspArgTrp-----ArgValArgLysGln 407
 Db 7157 CTCTTCATTTAAAGCTCTGTAGAAAGAAAGAGACAGACAGATACAGATTAAGAGAGA 7216
 QY 407 uAlaMetGlyLeuAlaGlnIleTyrrLys-----LysTyrrAlaLeuGlnSerAla 425
 Db 7217 ATCTAAACCTGAGTGGAGATGCTTCAGAAATCAGTTAAAGAGCTAAATGAGCAGTAGC 7276
 QY 425 agLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrr 445
 Db 7277 AGCCCTGTGTGTCACAGAAATTAAGAGCCACAGAACAGAGCTTACAGCCACCAAT 7336
 QY 445 rGlnAsnSerIleAspArgLeuLeuValGlnArgIlePheAlaGlnTyrrMetValPr 465
 Db 7337 AGAGAAAGACATCAGCTGAGAAATAGCATTAAGAGCTGAGAGCCCGC----- 7385
 QY 465 oHisAsnLeuGlnThrThrGluArgMetLysCysLeuTyrrLeuTyrrAlaThrLeuAs 485
 Db 7386 -----CTAGAACCTGATGAAGAAAGAGCAG----- 7409
 QY 485 pLeuAsnAlaValLysAlaLeuAsnGlnMetTrpLysCysGlnAsnLeuLeuArgHisGln 505
 Db 7410 -CTCTGTCTTACAAACACTGMAAGAAAGTACATCATGCAATTAATTAAGAGTAG 7468
 QY 505 nValLysAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVal 522
 Db 7469 AGTGAAGAACCTTGAAGAGAGCTAGATAGCCAGAGAACCAAGACATGCACTCT 7528
 QY 522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
 Db 7529 TGAGCGACAGAAATTCCAAAGAGAGAGTAGAGACCTTAAGAAATAGAAAGGATGAC 7588

QY 542 agLysPhe-----MetLysLysPhe 549
 Db 7589 CCAAAAGCTGAGAGCTCGGAATTAAGTGTGTCTATAGCTGAGAAAAAGAAATCT 7648
 QY 549 eThrGlnValLeuGlnAspAspGluLysIleArgLysGlnLeuGlnValLeuValSerPr 569
 Db 7649 GACAAATGATTTACAAAAGAGACAGAGCATATCTGATTAAGAAATTAATTAACA-- 7706
 QY 569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuLys 589
 Db 7707 -----TCATTTGAAAAATATTTTGCAAGAAAAAGACAAAGAAAGTACA 7750
 QY 589 nProLysGlnProThrAsnProPheLeuGlnMet-----IleLysPheLeuLe 605
 Db 7751 GATGAAGAAAAATCAACGACTGCCATGGAGTGTCTTAACACAAATTAAGAGCTCA 7810
 QY 605 uGluArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLysGlnVal 625
 Db 7811 TGAGAGAGTGGCAGCCCTGCATATGACCAAGAACCTGTAAAGCC-----AAAGAGA 7864
 QY 625 lAsnLysSerIleAspGlyThrAlaAspAspGlnAspGluGlyValProThrAspGlnAl 645
 Db 7865 GAATCTTAGT-----AGTCAAGTAGAGTGTCTTGAACCTGAGAAAGGC 7906
 QY 645 a---IleArgAlaGlyLeuGlnLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
 Db 7907 TCAGTTGTGTACAAAGCCTTGAAGGCCAA-----AATATATATATGTTT 7954
 QY 664 eHisSerAla-----GluThrPheGlnSerLeuLeuAlaCysLeuLysMet 679
 Db 7955 GCATCTTCAGTGAAGAGCCCTCATTCACAGAGTAAAGATGCGCAAGCAGAACTGAGAA 8014
 QY 679 tAspAspLysValAlaGlnAlaLeuGlnIlePheLysAsnThrGlySerLysIle 699
 Db 8015 GAAGCATGAGAAATCAGTAGAGCTGAAAAATCAATT----- 8051
 QY 699 eGluLysPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis-- 716
 Db 8052 ---CAAGACCAAGAGACCTGTCTTAACTGTCCAGGTCGAGAGAGACAGACCAACT 8107
 QY 717 ---LysSerLysGlyProProArgGlnAlaLysTyrrAlaIleHisCysIleHisAl 735
 Db 8108 TTGGAAGAGCAAACTTAAGAACTGAAATGTGACGGTGAATTTGAGCAGAAAGATGCA 8167
 QY 735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
 Db 8168 AGTGTACAAATCCAAAAATGCTTTCAGAGACACATTAAGAAAGTGTGACAGATCTTA 8227
 QY 755 uAspProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe 775
 Db 8228 C-----AAGAAATCTAGAGAAAT-----GACCTTGAAAT 8254
 QY 775 uLeuAlaProAspGlnPheAlaIleProThrLysSerTrpValAlaThrPheIleValL 795
 Db 8255 GACAAAAATGCACAAAATG-----TCCTTTGT----- 8282
 QY 795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
 Db 8283 -----GAAAAAGTAAACAAAAAGTACAGTGCAGAAAGGA 8311
 QY 815 pGluGlnValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
 Db 8312 AACTGAGTGCAGAGAGAAATGCATGAGATGGCAGAGAAACAGCAGAGCTCAAGAGA 8371
 QY 835 pLeuLeuGlnLysMetLysAsn-----AsnHisSerLys 845
 Db 8372 ACTGATGAGAGAAATTAAGCTAGCTGAGAGAGTTCAGATTAAGTGAAGAAATTA 8431
 QY 845 sSerGlyThrSerThrLeuAlaGlnLeuThr----- 855
 Db 8432 GAGCAGCAAAAGATCAATTTGAAGAGAGCTCAGACCTAGAAATAGTGAATTTGAAGAGAGCT 8491

OY		856	ThrlleuLeuIserSerpgLysapLeuThrGlnGlnGlySerLysIleSerLysProAspMe	875
Dd		8492	AGATTGCATGGACAAAGAC---CAGGTGGAAAAGGAAGAAAGTG--AAGAGGGAAT	8545
OY		875	tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy	895
Dd		8546	AGCTGAATATCAGCTACGGCTTCATGAAAGCTGAAGAAGAA-----	8585
OY		895	rHisGluIleIlethrLeuGlnGlnTyrGlnLeuCysAlaLeuAlaIleasnspGluCy	915
Dd		8586	-CACCAAGCTTTGGCTTTGGAC-----ACAAACAACA	8617
OY		915	sTyrGlnValArg---GlnValPheAlaGlnLysLeu-----HisLysGlyLeu	931
Dd		8618	GTAAGCACTAACAAATCCAGACATATCCAGAGAAAAATTGACTCTTAAGACGAATGTCTCAG	8677
OY		931	rArgLeuArgLeuProLeuGlnTyrMetAlaIleCys-----AlaLe	945
Dd		8678	TTCACAGAAAGCTGGAGATAGACGCTTTAAAGTCAAGTAAAGACAGCTCATATTCATT	8737
OY		945	wCysAlaLysasproValLysGluAlaGlaAlaHisAlaArgGlnCysLeuValLysAs	965
Dd		8738	GAAAGCTCTACTCAGATTGGTGAAGAAATTGAAGAAACCAGATGGCAATCTAAATA	8797
OY		965	nIleasnValArgArgGluTyrLeuLysGlnHisAlaAlaValArgGlnLysLeuLeuSe	985
Dd		8798	TGTAAAT-----CAGTTGAGAGAGAAATGACGTGCCCGGGGAAATGAA	8845
OY		985	rLeuLeuProGlnTyrValValProTyrThrIleHisLeuLeuAlaHisAsproAspTy	1005
Dd		8846	GTGTGTG-----	8852
OY		1005	rValLysValalnspIleGlnGlnLeuLysAspValLysGlnCysLeuTrpPheVal	1025
Dd		8853	-----ATCAAATCCCTGTAACACGCTGGAAGAGAAAGAGATCTGCAGAAAGAAGCT	8905
OY		1025	wGluLeuLeuMetAlaLysasnGluAsnAsnSerHisalaphelleArg-----	1041
Dd		8906	CTCTCACTTCAAGCTGCACAGAGAGAACACAAAACAGGTACTGTATGATRCCAAAGCT	8965
OY		1042	---LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaly	1060
Dd		8966	CGATGAATTAACACTGATGATCAAGAAGCTGAAGAAACCTGTGAAGAAAAACCAAGGA	9025
OY		1060	sMetasnGluLysLeuTyrThrValCysAspValAlaMetasnIleileMetSerLysSe	1080
Dd		9026	GGCAGATGAATACCTTGGATTAAGTACTGT-----TCCTTGCTTATMAACCATGA	9073
OY		1080	rThrThrTyrSerLeuGlnUserProLysAsp-----ProValLeuProAlaArgPhePh	1098
Dd		9074	A-----AAGTAGAGAAAGCTMAAGAAATGTAGAGACACACAGTGGCCACTCTGTG	9124
OY		1098	eThrGlnProaspLysasnPheSerAsnThrLysAsnTyrLeuProProGlnMetLysSe	1118
Dd		9125	TTCACAGCAATCTAAACAAGATTCGCCAGGGGTCTCCTTGTAGTGCAGTTGT----	9179
OY		1118	rPhePheThrProGlnLysProLysThrThrAsnValleuGlnAlaValasnLysProLe	1138
Dd		9180	-----CCAGGACCATCTCCAATCCCTCTGTACTGAAGAAGAGGT	9220
OY		1138	uSerSerAlaGlnLysGlnSerGlnThrLysSerSerArgmetGluThrValSerAsnAl	1158
Dd		9221	ATCATCTTGGCAAAATAAAGCTTCACGCCAAGGCCAAAGATCCAGTGGAATATGGAGAA	9280
OY		1158	aSerSerSerAsnProSerSerProGlnArgIleLysGlnrGluLeuaspserserGln	1178
Dd		9281	TGGTAGAGGACCAACACTGCTACCCA-----	9308
OY		1178	uMetAspHisSerGluasnGlnUaspTyrThrMetSerSerProLeuProGlnLysLysSe	1198
Dd		9309	-----GAGAGCTTTTCT-----AAAAAAG	9328
OY		1198	rAspLysrArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArgIleArgLy	1218

```

Db      9329  CAGAAAAGCATGATGAGTGGTATATCAACCTGCAGAA----- 9365
           |||      |||      ::|      ::|||
Oy      1218  slvsthrprovalthrsluglugsleuylmetaspseuthrlysluYa 1238
           |||      |||||      |||      ::|      |||
Db      9366  -GACACGGAAGTACTGAGTTTGAGCCAGAG---GCACTTCAGAAAGTGTAAAGAAAG 9421
           |||      |||||      |||      ::|      |||
Oy      1238  lGlnglugslnlyspolysglY-----serGlnarGserArlylsarGlylthsthrAl 1256
           |||      |||      |||      |||      |||
Db      9422  GPTTGCAGATCCCGACAGAAAGACTAGCCCATATATCTCGTGAAGAACCAATGCGC 9481
           |||      |||      |||      |||      |||
Oy      1256  aserGlnseraspGlnuglugslntrpProglugslyArGleuLysgluspileueu-- 1275
           |||      ::|      ::|      |||      |||
Db      9482  AACTCGGACGACCCCGCCGCTGCGTGCAGAAAGTTAGCCGTATCCCACTGAGTCTCGG 9541
           |||      ::|      ::|      |||      |||
Oy      1276  ---GluasngluspGlnuglugslnsSerProPolysLysglYlylsarGlylarGpropR 1294
           |||||      |||      |||      |||      |||
Db      9542  CAAAGAAATCTTCAGAGTCTCCAAACCAACAGCTGTGGCAGCAGATCACAAAGGT 9601
           |||      |||      |||      |||      |||
Oy      1294  olys-----ProleuGlyGlyGlythProlyslugslnuProthrmetyl 1309
           |||      |||      |||      |||      |||
Db      9602  CAAAGTTGCTCAGCGGAGCCAGATGATTCAGGCACCATCTCCGAAACCCACAGAA 9661
           |||      |||      |||      |||      |||
Oy      1309  s-----ThrseryslsGlyse 1315
           |      |||      |||      |||      |||
Db      9662  ATCGCTCCGATCAATATCTCTCTGAGAGAAAGTCGACTGACAGACCCCAAGAGAGGCT 9721
           |||      |||      |||      |||      |||
Oy      1315  rlyslsLysSersGly-----ProProAlaProGlu 1325
           ::|      |||      |||      |||||      |||
Db      9722  GAGGGTCAAGCGAGCGCCGACTTGTCCACGCCCCCAAG 9758
           |||      |||      |||      |||      |||

RESULT 7
; US-08-328-254-5
; Sequence 5, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hua
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

FEATURE:
NAME/KEY: CDS
LOCATION: 544..7990
US-08-328-254-5

Alignment Scores:

Pred. No.: 1.36e-07 Length: 8789
Score: 197.00 Matches: 280
Percent Similarity: 35.77% Conservative: 248
Best Local Similarity: 18.97% Mismatches: 604
Query Match: 2.74% Indels: 345
DB: 1 Gaps: 62

US-09-512-581b-2 (1-1391) x US-08-328-254-5 (1-8789)

```

OY 19 ValLysGluLLeSerAspLysLLe-----SerLysGluLLe 30
DB 4057 ATGAAGAATTAGCTCAAACTCCATTACAGAGGTACAACTAATGACCAAAATTGAA 4116
OY 31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
DB 4117 GCATGCATGAAATTGGAAAAAATGTTGGGAAGTAAAGAAAGAAACTCAGATTAAAGT 4176
OY 51 GluLysGluLLeuLys-----56
DB 4177 GAAAAATTGGAATATTTCTTGTGATCACCAGAGTTACTCCAGAGTAGAAACTTCT 4236
OY 57 -----LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLys 74
DB 4237 GAAGGCCCTCAATTCCTGATTGAAATGCATGCAGAT-----AAATCATCAGGTGA 4287
OY 75 AspValArgLeuLeuValAlaCysCysLeuAlaAspLLePheArgLLeTyrAlaProGlu 94
DB 4288 GATATTTGAGATATATGTGCG-----4308
OY 95 AlaProTyrThrSerProAspLysLeuLysAspLLePheMetPheLLeThrArgLLeu 114
DB 4309 -----AAGGTGAATGACAGCTGGAAGAGAGATTTCTTGATGTGAAATGAGCTG 4359
OY 115 LysGluLLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLLeuGluAsn 134
DB 4360 AGTAGATCAGATCGAGAGAAAGCTAGCATTCAGCTGAAGCCCTTACTCGAGAGCTGAC 4419
OY 135 IleAlaTrpValLysSerTyrAsnLLeCysPheGluLeuGluAspSerAsn-----Glu 152
DB 4420 TTAGAGTAGATTCAACAGAGAGACGTATTTAGAAAAAGACATGAAATTAAGCAGAG 4479
OY 153 LLePheThrGluLeuTyrArgThrLeuPheSerValLLeAsnAsnGlyHisAsn-----170
DB 4480 GTTATTTGTCCTTGAAGAAGAACTC---TCAGTGTGCACAGAGAGAGAAACCGACTT 4536
OY 170 -----170
DB 4537 CGTGAGAAATTAGATCTATGTCAAAAAACCAACGCGCATGATCACTGTCGAAAAA 4596
OY 171 -----GlnLysValHisMetHisMetValAspLLeuMetSerSerLLe---184
DB 4597 ATGAAGAGAAAAACAGACAGCTTACGTCTCATCAAGTGAAGTGTCTCATTGCATTGAG 4656
OY 185 LLeCysGluLLeAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
DB 4657 GTGCGACAGCGCAGAGTAAGAAAGAGGAGAACTCTTCAGACTTCTCTCGATGTG 4716
OY 204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLLeuAlaLysAlaLeu-----221
DB 4717 -----AGTGACGTGTTAAAGACAAACTATCTCCAGGAAACCTCAGAGAGT 4764
OY 222 LeuLysArgThrAlaGlnAlaLLeGluProTyrLLeThrPhePheAsnGlnValLeu 241
DB 4765 TTGGAAGAAAGCATCAGAGGCACTGTCTTGACAAATGTAGCTGGAAACCAATTTGCA 4824
OY 242 MetLeuGlyLys-----ThrSerLLeSerAspLLeuSerGlnHisVal 255

```

```

DB 4825 CAACTGAATTAAGAGAAAGAAATTGCTGTCAAGAAATCGAAAGCCCTGACGACACTG 4884
OY 256 PheAspLLeLLeuGluLeuTyrAsnLLeAspSerHisLLeuLeuLeuSerValLeuPro 275
DB 4885 AGTGAATCAGATTATGAAGAGCTGATCTTCCAAAGCCCTTGAGCCGCTGCTGGAG 4944
OY 276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu-----ArgLeu 290
DB 4945 AAAGGTGAGTTTGCATTAGCGGTAGCTCAACACAGAGAGAAATGCTACATCAGTGAGAA 5004
OY 291 GlnValValLysLLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGln 310
DB 5005 GGCATCGAGAACTGAGAGTTGCGATTCAGGCGCATGAAAGAAAGACACCTGCATCCCA 5064
OY 310 AsnLysProLeuThrProGlnCysTyrLeuGlyArgPheAsnAspLLeHisValProLLeArg 330
DB 5065 GAGAAAC-----TGAAAGAACCGCAGCGGAGAGATG-----ATTCACTTAAGATAAA 5112
OY 330 GLeuLLeuCysValLysPheAlaSerHisCys-----Le 341
DB 5113 GTTGGAACCTTGAAAGGGAATTGCGATGTCAGAGAAACCAAGAGCTGATGCTTCTT 5172
OY 341 MetAsnHisProAspLLeuAlaLysAspLLeThrGluTyrLeuLysValArgSerHisAs 361
DB 5173 GATG-----CCGAGAAATTCGAAAGCAGAA-GTAGAGACTGTAAAAACACAAATGAGAGA 5225
OY 361 ProGluGluAlaLLeArgHisAspValLLeValSerLLeValThrAlaAlaLysLysAs 381
DB 5226 GATGCCCGAAGCCGGAAGATTTGGAATTGAGCTTGCACGTTAAGTGTGAAAAAGA 5285
OY 381 PLeuLeuValAsn-----AspHisLLeu 390
DB 5286 AAATCTGCAAAACAAATACAGAAAAACAAGGTGACGTTGCACAGCTAGACAACTTACT 5345
OY 390 AsnPheValArgLLeuArgThrLeuAspLysArgTyr-----ArgValArgLysGln 407
DB 5346 CTCTTCATTATTAAGTGTGTAGAAAGAAAGAGACAGACAGATACAGATCAAAAGAGA 5405
OY 407 ValAlaMetLLeLLeuAlaGlnLLeTyrLys-----LysTyrAlaLeuGlnSerAlaAl 425
DB 5406 ATCTAAACCTCAGTGGAGAGATGCTTCAGATCAGTTAAAGAGACTAAATGAGCGCATGAC 5465
OY 425 AgLysAspAlaAlaLysGlnLLeAlaTrpLLeLysAspLysLeuLeuHisLLeTyrTy 445
DB 5466 AGCCTGTGTGTGTGACCAAGAAATTAATGAGCCACAGACAGAGCTAGAACCCACAT 5525
OY 445 GlnAsnSerLLeAspAspArgLeuLeuValGluArgLLePheAlaGlnTyrMetValPr 465
DB 5526 AGAGGAAGAGCATCAGCTAGGAAATATGATTTGAAAGCTGAGAGCCGCGC-----5574
OY 465 OhHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLLeTyrAlaThrLeuAs 485
DB 5575 -----CTAGAACCTGATGAAGAAAGAACAG-----5598
OY 485 PLeuAsnAlaValLysAlaLeuAsnGlnLLeMetTrpLysCysGlnAsnLeuLeuArgHisGln 505
DB 5599 -CTGTGTCTTTACACAACTGAAGAGAAAGAGAGACATCATGATTTACTTAAAGGATG 5657
OY 505 nValLysAspLeu-----LeuAspLLeLLeLysGlnProLysThrAspAlaSerVal 522
DB 5658 AGTGAGAAACCTTGAAAGAGAGCTAGAGATGAGCCAGCAACCAAGAGCATGAGAGCTCT 5717
OY 522 LLeValAlaLLePheSerLysValMetValLLeThrArgAsnLeuProAspProGlyLysAl 542
DB 5718 TGAGGCAAGAGATTCCAAAGAGAGATGAGAGACCTTAAGCAAAAAATAGAGAGATGAC 5777
OY 542 agLAspPhe-----MetLysLysPhe 549
DB 5778 CCAAGAGCTGAGAGCTCGAATTAGATGTTGTTACTATAGCTACAGAAAAAGAAATCT 5837
OY 549 eThrGlnValLeuGluAspAspGluLysLLeArgLysGlnLeuGluValLeuValSerPr 569
DB 5838 GACAAATGCAATTACAAAAAGAGCAAGAGCAATATCTGCAATTAAGAAATTAATTTCA--5895

```

QY 569 orthCysSerCysLysGlnIleGluCysValArgGluIleThrLysLysLeuGlyAs 589
 Db 5896 -----TCATTGMAAATATTGCAAGAAAGAGCAAGAGAAATGACA 5939
 QY 589 nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeu 605
 Db 5940 GATGAAGAAAGAAATCAAGACCTCCATGAGATGCTTCAACAACTAAAGAGCTCAA 5999
 QY 605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
 Db 6000 TGAAGAGTGGACCCCTGCATATATGACCAAGAGCTGTAAAGCC-----AAGAGCA 6053
 QY 625 IasnLysSerIleAspGlyThrAlaAspAspGluAspGluIleValProThrAspGlnAl 645
 Db 6054 GAATCTTACT-----ACTCAAGTAGAGTGTCTTCACTGACTGAGAGGC 6095
 QY 645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
 Db 6096 TCAGTTGCTTCAAGGCCCTTGATGAGGCCAAA-----AATTAATTATATTGTTT 6143
 QY 664 eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
 Db 6144 GCATCTTCAGTAAATGGCCCTCATTCAGAGTAGAGAGATGCCAGCAAACTGGAGAA 6203
 QY 679 tAspAspGluLysValAlaGluAlaIleLeuGlnIlePheLysAsnThrLysSerLysI 699
 Db 6204 GAAGGATGAAGAATCAGTAGACTGAAATAATCAATT----- 6240
 QY 699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis-- 716
 Db 6241 ----CAAGACCAAGAGAGCCTGCTCTTAACCTGCCAGCTGAGAGAGAGCAACCACT 6296
 QY 717 ----LysSerLysLysGlyProProArgGlnAlaLysTyraIleHisCysIleHisAl 735
 Db 6297 TTGGAAGAGCAAACTTACACTGAGAAATCTGCAGCTGGAATGGAGCAAGATCCA 6356
 QY 735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
 Db 6357 AGTGCTACATCCAAAATGCTCTTTCAGAGACACTTAGAAGTGCTGAGAGTTCTTA 6416
 QY 755 uAspProSerAsnLeuGlnHisIleuIleThrProLeuValThrIleGlyHisIleAlaLe 775
 Db 6417 C-----AAGCAATCTAGAGAT-----GAGCTTGATT 6443
 QY 775 uIleAlaProAspGluPheAlaIleAlaProTrpLysSerTrpValAlaThrPheIleVal 795
 Db 6444 GACAAAATGCAAAAATG-----TCTTTGTT----- 6471
 QY 795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
 Db 6472 -----GAAAAAGTAACAAAATGACTGCAAAAGA 6500
 QY 815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetIleValArgTr 835
 Db 6501 AACTGTAGCTGCAGAGGAAATGATGATGAGCAACAGAAAAAGCAGAGCTGCAAGAGA 6560
 QY 835 pIleuLeuGlyMetLysAsn-----AsnHisSerLys 845
 Db 6561 ACTCAGTGGAGAAAATATAGCTAGCTGAGAGCTTCACTGTGGAAGAAAATATA 6620
 QY 845 sSerGlyThrSerThrLeuArgLeuLeuThr----- 855
 Db 6621 GACACACAAAGATCAATTGAAAGAGCTCACACTAGAAAATAGTAATTGAAGAAGACCT 6680
 QY 856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
 Db 6681 AGATTGCATGCAGAAAGAC---CAGGTGAAAAGAGGAAAGAGTG---AGAGAGCAAT 6734
 QY 875 tSerArgLeuArgLeuAlaIleAlaGlySerAlaIleValLysLeuAlaGlnIleProCysTy 895
 Db 6735 AGCTGAATATCAGCTACGGCTTCATGAGAGCTGAAAAGAAA----- 6774

QY 895 rHisGluIleIleThrLeuGluGlnIleGlnLysCysAlaLeuAlaIleAsnAspGluC 915
 Db 6775 -CACCAAGCTTGTGTTGAC-----ACAAACAACA 6806
 QY 915 sTyrglnValArg---GlnValPheAlaGlnLysLeu-----HisLysGlyLeu 931
 Db 6807 GTATGAAGTGAATAATCCAGACATACCGAGAAATGACTCTTAAGAAAGAAATGCTCAG 6866
 QY 931 rArgLeuArgLeuProLeuGlnIleThrMetAlaIleCys-----AlaLe 945
 Db 6867 TTCACAGAACTGGAGATACCTTTAAAGTCTAGTAAAGAGAGCTCAATATTCATT 6926
 QY 945 uCysAlaLysAspProValLysGluArgAlaHisAlaArgGlnCysLeuValLysAs 965
 Db 6927 GAAAGCTACTACACAGATTTTGGAAAGATTAAGAAAACCAAGATGACATTAATAATA 6986
 QY 965 nIleAsnValArgArgGluIleThrLeuLysGlnHisAlaIleValSerGluLysLeuLeu 985
 Db 6987 TGTAAAT-----CAGTTGAAGAGAAATGAACGTGCCAGGGAATAATGAA 7034
 QY 985 rLeuLeuProGluTyraValValProTyThrIleHisLeuLeuAlaHisAspProAspTy 1005
 Db 7035 GTTGTTC----- 7041
 QY 1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuThrPheValLe 1025
 Db 7042 -----ATCAAAATCCTGTAACACAGCTGGAAGAGAAAGAGATCTGCAAGAAAGCT 7094
 QY 1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg----- 1041
 Db 7095 CTCTCAACTTCAAGCTGCACAGAGAGAGAAACAGTACTGTTATGATACCAAGGT 7154
 QY 1042 ----LysMetValGluAsnIleLysGlnThrThrLysAspAlaGlnGlyProAspAspAla 1060
 Db 7155 CGATGAATTTACACACTGAGTCCAAAGACTGAAAGAACTCTTGAAAGAAAAACAAGCA 7214
 QY 1060 sMetAsnGluLysLeuTyThrValCysAspValAlaMetAsnIleLeuMetSerLysSe 1080
 Db 7215 GGCAGATGAACTACTGTGATAAGTACTGT-----TCTTGCTTAAAGCATGA 7262
 QY 1080 rThrThrTySerLeuGlnSerProLysAsp-----ProValLeuProAlaArgPhePh 1098
 Db 7263 A-----AAGTTAGAGAAAGCTTAAGAGATGTAGAGACACAAGATGCCATCTGTG 7313
 QY 1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyLeuProProGluMetLysSe 1118
 Db 7314 TTCACAGCAATCTTAACACAAATTCGCCAGGGTCTCTTCTCTAGGTCACATTTGT----- 7368
 QY 1118 rPhePheThrProGlyLysProLysThrThrAsnValIleuGlyAlaValAsnLysProLe 1138
 Db 7369 -----CCAGGACCATCTCCAAATCCCTTCTGTACTGAAAAGAGTT 7409
 QY 1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerSnaI 1158
 Db 7410 ATCATCTGGCCAAATAAAGCTTCAGGACAGGCAAGAAAGTCCAGTGAATATGAGAGAA 7469
 QY 1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerG 1178
 Db 7470 TGGTGAAGAGCAACACCTCTACCCAGAGAGCTTTTCTAAAAAAGCAAGAAAGCAGT 7529
 QY 1178 uMetAsp-----HisSerGluAsnGluAspTyThrThrMetSerSerProLeuProG 1195
 Db 7530 CATGAGGATTCACCCCTGCAGAGACAGCAAGAGTACTGAGTTGAGCAAGAGGACT 7589
 QY 1195 yLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlnIleProAr 1215
 Db 7590 TCCAGAGAGTTGTAAGAAAGGGTTTCTGCATCCGACAGAGAAAGACTAGCCCATATAT 7649
 QY 1215 gGlyArgGlyLysThrProValThrGluGlnIle-----GluLysLeuG 1230
 Db 7650 CTCGCAAGAACACCATGCACTGGACAGCAGCCCGCTGCTGACAGAAAGTTAGC 7709
 QY 1230 yMetAspAspLeuThr-----LysLeuValGlnGluGlnLysPro--LysG 1245

QY	339	CysLeu-----	-----	metSnHsProAsp-----	346
Db	4387	CGTTTGGAGTGAAGACGTAAGACCTGACGTCATACAGACGACCTGGACAACTGAT			4448
QY	347	-----LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluAla			365
Db	4447	GAACCTGATATAAACTTTTGAGTGA-----GTCAAACTTAAGTGGAAACTGTGATTAAAC			4503
QY	365	AlaLeuArgHisAspVal-----			371
Db	4504	AGGAAGACATATGTCACAGAAACAGCAACGACAAACCCAAAGGATGATGAGCACT			4563
QY	371	evalSerLe-----	-----	ValThrAlaAlaLysLysAs	381
Db	4564	GACCTTCCTCGAGGCTCTTTACAAATGACCTGGGCGCAGAGGTGACAGAAAGAAACAGCA			4623
QY	381	pIleLeuLeuValAsnAspHisLeuLeuAsnPhenValArgGluAspThrLeuAspLysArg			401
Db	4624	TCTGGAAGACGATACAGCTGTTGGCCCGGAAATGAAAGAAAGCGCTGTTCTCTCTGCA			4683
QY	401	gTPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAlaLe			421
Db	4684	ATGGCTTCTCGTACTGAA-----	ACGTAATTTGGACAGAAAGTCCACATTC		4728
QY	421	uGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleLeaTrpIleLysAspLysLeuLe			441
Db	4729	AGAAAGTGTGCTTGGT---GACCTTGATACAGAAATTTCCGCGCTAAATAATTTCTGAA			4785
QY	441	uHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaG			461
Db	4786	GGATGTGGAAAAAGAAAAAGCTGATTTAAATACCATCACAAGAGTACTGCTGCCCTGCA			4845
QY	461	nTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTy			481
Db	4846	AAACTTGTATGGGCGACGTGAGCCATTTTGAAGAGAGAGCGCTGCGCTC-----			4894
QY	481	ralaThrLeuAspLeuAsnAla-----	ValLysAlaLeuAsnGluMetTrpLys		497
Db	4895	-----CTTAACGCTGGGTGAGCCGAGATTGCTACTGAGCTGAAGATTGG--			4939
QY	497	scYsGlnAsnLeuLeuArgHis-----	ValLysAlaLeuAsnGluMetTrpLys		504
Db	4940	-TGCATATCCTTGTATGAAACCATCAGACCACTAGCAATATTGTATGGAAACGTGGCTCA			4998
QY	505	-----GlnValLysAspLeuLeuAspLeuLysGlnProLysTh			518
Db	4999	CATAGTACCTGGCTTTATCAAGCTGAAGCTCTATTGTGTAATTAAGAAAGAAACCAAC			5058
QY	518	rasPalaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAs			538
Db	5059	A-----	-----	-----	5059
QY	538	pProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp----			556
Db	5060	-----AGTAAACAGAGAAATTTGTACACGCTTTAGTATCTAGCTGGATGATGCCAACCT			5115
QY	557	-----GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLys			574
Db	5116	CCAGGTGAATGATGCCGCGATCAACGCCCTTATTGTGTAATGGCCGTGGAAGCTCAG			5175
QY	574	sgLnaIaGluGlyCysValArgGluIleThrLys-----	LysLeuGlyAsnTrp		590
Db	5176	C-----AGGAGCTTGTGAACAAAGTCAAAAGTTCGTAATTTGCTACGAGAACCATTT			5217
QY	590	olysGlnProThrAsnProPheLeuGluMetLleLysPheLeuLeu-----	GluArgIle		608
Db	5218	CTTTGAAGAAAGGTGCTCAACATATCAAAAGTGCMAAATTTGCTAATTTGCTACGAGAACATTT			5277
QY	608	alaIProValHisIleAspThrGluSer-----	-----	-----	617
Db	5278	ATACCAATGTTTGGTACACACTGAACATTTGAAACATGCTGCTTCTCTGACTTGA			5337
QY	618	-----IleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh	632		

Db	5338	AAAAATTAGAAAATGACATAGAAAATATGTTAAATTTGTGGAAAAAACCTGGAAATCCAG	5397
Qy	632	ralaspaspglu-----Asp[IGL]yvalProthrAsp[Inal]leArg[al]	649
		:::	
Db	5398	TCATCAACATGTAAGAAGATGCATGAGACAGAGCCACAGATTGAGAAAGTTCTACAAAGAG	5455
Qy	649	yleu[IGL]leuleu[ys]val[leu]SerPheThrHisPro[le]SerPheHisSer[al]gluTh	669
		:::	
Db	5458	AGAAAGAAATGTTACAT-----CAACCTGAGACAGATATATAAAAAGAAA	5502
Qy	669	rPhe[le]Ser[leu]leu[al]ac[ys]leu[ly]Met[asp]p[lu]ys[al]al[glu]al[al]le	689
Db	5503	GATCCGTTTGCAATTAATTACTTTTGCAATCTAGATATACAAACAAAATTAAGCAATCCAT	5562
Qy	689	ugl[In]lePhe[lys]as[thr]ngl[ys]er[il]egl[lu]asPheProHis[le]Arg[ser]	709
		:::	
Db	5563	TCAA-----CAGAGAAAATGGGTCACTGCTCTTGGAATTAGATCATC	5607
Qy	709	aleu[leu]ProVal[leu]His[His]Lys[Ser]Lys[gl]yPro[Arg]Inal[al]ys[tyr]	729
Db	5608	ACTCTTCTCT-----ACAGATTATCT	5628
Qy	729	alleHis-----	731
Db	5629	GGTGAATAATTAAACAAATTTACTTTGCATGATGATGTGAATATATGCTTAATGTTCC	5688
Qy	732	-----Cys[le]His[al]lePhe[Ser]Ser[ys]gluThr[In]Phe[al]glu[In]	747
Db	5689	AGAGCTCAACACTGCTATTTACGAAGACTTCTCTTTTACGAAAGACTCTCGAAGAAAT	5748
Qy	747	ePhe[glu]Pro[leu]His[lys]-----	753
		:::	
Db	5749	CAAGAGCAACCTGAGCAAACTTGAGAGCAGATTGCAGTCAATTCATGAAAAACACCAGA	5808
Qy	754	-----Ser[leu]asPro[Ser]as[In]leu[glu]-----His[leu]le-----	764
Db	5809	TGTCATCCTTGAAAGCCTGTGACCTGGAAGCCATTCAGATCAGATATCACTTACTCACT	5866
Qy	765	-----ThrPro[leu]Val-----Thr[il]egl[His]le[al]leu[al]leu[al]	777
Db	5869	GAATGCAAAATGGGACAGATTAATAGATAGTACAGTGTATCGAAGAGTTGTTTG	5922
Qy	777	laPro[asp]InPhe[al]la[al]aPro[tyr]Lys[Ser]tyr[Val]al[al]ThrPhe[le]Val[lys]asPl	797
		:::	
Db	5925	-----ACAGGCAATGGAAGATGAGAGACAGTTCATTCATTCAGACC-----	5963
Qy	797	euleu[Met]as[par]glu[leu]Pro[gl]y[ys]Thr[Thr]Lys[leu]Tr-----Val[Pro]asp	816
Db	5964	--TTAATACCTCACACAGTGAATACAGAGGCTGGAAGATTAAGTGGTTGATACCTGTGC	6021
Qy	816	lu[glu]Val[Ser]Pro[glu]Thr[Met]Val[ys]Il[glu]Inal[il]le[ys]Met[Met]Val[Arg]PL	836
		:::	
Db	6022	TCCAGGTGGCAGCCCTGAGACTTAGAGAAAGCCAGATATATACAGCAGAACTTGAGGTGG	6081
Qy	836	euleu[glu]Met[lys]as[His]Ser[tyr]Lys[Ser]glu[Thr]SerThr[leu]Arg[leu]ThrT	856
		:::	
Db	6082	C-ATC-----AGCAGCCACAGCCACAGTTTGCAGCA-----	6112
Qy	856	hr[le]leu[His]Ser[as]P[leu]ys[leu]Thr[glu]In[glu]Lys[il]le[Ser]Lys[Pro]as[Met]	876
		:::	
Db	6113	-----CTTAACCGAAGTGGGATGGATGTGCAG--AACTCTCCAGAGCAATGCA	6164
Qy	876	etArg[leu]-----Arg[leu]al[al]glu[Ser]al[le]y	887
Db	6165	GCCTTCTTAAGAAAACCTGCGAGGTTTAAACCAACCTGGGAGTCAATCTTGCACGAAG	6224
Qy	887	al[ys]leu[al]glu[In]Pro[cy]s[tyr]His[glu]le[le]Thr[leu]glu[In]tyr[In]---	905
Db	6225	TGAAGGATAGGCGCAAGGCTAAAGAGAAAGTAAACGACGTGATGATGAAAGGCTC	6284
Qy	906	-----Leu[cy]s[al]leu[al]le[as]as[Plu]cy[tyr]glu[Val]Arg[In]v	921
		:::	

```

Db      6285 AGCTAGATGAGATTAATCTGTTGGTTAAACAAGAGCTGACGATCTATGCAAAAGAGATCA- 6343
QY      921 aIhPhealaglnlyserhslslysglyleuSerArgleuArgleuProleuGluTyrMetA 941
Db      6344 -----ACCAACGAAATGGAGAAAACCTGCAGAAATTAAGA----- 6379
QY      941 lailecysAlaLeucysAlaLysAspProValIysGluArgArgAlaHisAlaArgLnc 961
Db      6380 -----GACTTAACCTCAAGAAATGGAAGACACTGCTGAAAAA- 6415
QY      961 yslLeuValIysAsnIleAsnValArgArgGluTyrleuIysGlnHisAlaAlaValSerg 981
Db      6416 -----CTCAAAATGGCTGAAT-----AGAACTGAATTTGGAGATG-----CTTTGAG 6455
QY      981 luIysLeuLeuSerleuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
Db      6456 ATAAAGCTGTGAGTTTA----- 6472
QY      1001 lAspProAspTyrValIysValGlnAspIleGluIleuIysAspVal----- 1017
Db      6473 -----CTGGAAGAGGATAAATTTCA-----GAAAGCTTAAGAGCTGAATATGACAT 6521
QY      1018 -----LysGluCysleuTTPheValLeuG 1026
Db      6522 GCAATAGATTTTGACAGAGAGTGCCTACCCCTGAGAGAAATGCATC----- 6568
QY      1026 luIleuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgIysMetValGluA 1046
Db      6569 -----CAGAGAGCCAGTT 6581
QY      1046 snIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluIysLeuT 1066
Db      6582 CTGTTTCACAGACAAAGATGCTGCTGCATCTAATGTCACAAAGGCG----- 6628
QY      1066 yrfThValCysAspValAlaMetAsnIleIleMetSerIysSerThrTyrTyrSerLeuG 1086
Db      6629 --GTCTAGATATCATCTGCTCAGATATTCCTGCTCAGTCTCATGCTACTGTAAGAAATTT 6686
QY      1086 luSerProLysAspProValIeuProAlaArgPhePheThrGlnProAspLysAsnPhes 1106
Db      6687 CAATTCTGCTGAT-----CTGTATTAATAACTATAA 6716
QY      1106 erAsnThrLysAsnTyrLeu-----ProGluMetLysSer---PhePheThrP 1122
Db      6717 CAGAACTAGCCACACGGCTGATTAATGACACAGATGCTGAAGTCCACATTTGCACAG 6776
QY      1122 roGlyLys-----ProLysThrThrAsnValIleuGlyAlaValAsnLysProL 1138
Db      6777 TTGGGAGATGAGAAAGATCATTAAGACCGTTTCCGAAATTAACAAGCGTCACT 6836
QY      1138 euSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnA 1158
Db      6837 TAGAACAGCGCCATCCTCAGCTGATTAATGTTTATTCATTTGACAGAAATTTGAAAAA 6896
QY      1158 laSerSerSerSerAsnProSerSerPro-----GlyArgIleLysGlyA 1173
Db      6897 AAGCTTCAGTTCATATGAAACAGCAATTAACAGAAAAATTGCAAGGGCTCAAGAAC 6956
QY      1173 rGleuAspSerSer-----GluMetAspHisSerGluAsnGluAspTyrThrM 1189
Db      6957 AGTGAGATGGCACCCAGCATGGCGTTGAGCTTAAGACAGCAGCAGCTTGAGACATGATTA 7016
QY      1189 etSerSerProleuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgS 1209
Db      7017 TTGACAGTCTT-----CAGTGAGATGACCATTAAGGAGAGAGATGAAGAACATGATGA 7067
QY      1209 erGluLeuGluLysPro-----ArgGlyArgLysLysThrProValThrGluG 1225
Db      7068 GAAATATATGAGGCTGACACTATATTTTCACAGCAAGCCGAGCGGATCCACTACCAAAAC 7127
QY      1235 ln-----GluGluLysLeuGlyMetAspAsp----- 1233
Db      7128 AAATTTCTGATTAACCAATACTGTTCAAGAACTGGGTCTGAGATGATGATGTCATGG 7187

```

```

QY      1234 -----LeuThrLysLeuValGlnGluIleuLysProLysGlySerGlnArgS 1249
Db      7188 CGTTGATTAACGCTCTCGCAAGAACTCTGGAGAAATAT-----GGAGTGAATGACA 7238
QY      1249 erArgLysArgGlyHisThrAlaSerGluSerAspGluIleuInrProGluGluLysA 1269
Db      7239 CAGAGATGTGAAGAAGAAC--ACAGAGTACTTAAACATCATG--ATCA 7286
QY      1269 rGleuLysGluAspIleLeuGluAsnGluAspGluIleAsnSerProProLysLys-Gly 1288
Db      7287 ATCTCAACCAAAAGTAT-----GCTGACAGACAGAACCCCTTGAGGCTGATGA 7337
QY      1289 LysArgGlyArgProProLysProleuGlyGlyLysThrProLysGluIleuProThrMet 1308
Db      7338 GACAGGTGCAGGCTCTCTCCACAGATC-----TGG 7367
QY      1309 LysThrSerLysGlySerLysLysLys 1318
Db      7368 AAACCTCTCTGAAGTGATGATCAAGAAAGAG 7397

RESULT 9
US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

Alignment Scores:
Pred. No.: 5 9e-06 Length: 30549
Score: 190.00 Matches: 276
Percent Similarity: 35.29% Conservative: 235
Best Local Similarity: 19.06% Mismatches: 523
Query Match: 2.64% Indels: 414
Gaps: 67

US-09-512-581b-2 (1-1391) x US-09-134-001C-322 (1-30549)
QY      23 SerAspLysIleSerLysGluIleuValArgArgLeuLysMetValIysThrPhe 42
Db      26830 AATGATGAATATTAATGAATATTAATTCATATGACAAATTCATGAAATGCTGAATCAGCA 26889
QY      43 MetAspMetAspGluAspSerGluIleuLysGluLeuTyrLeuAsnLeuAlaLeuHis 62
Db      26890 TTTGATTTAAATGAATGACATGAATCAATCTTAAGAGATGAACCTTAATATCAAGCTTTTCT 26949
QY      63 LeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCys 82
Db      26950 GTTCAAGCAAGCTCAAAATTAATTAATTCAGATGAAGATTTAAACAAACAATTCACCAT 27009
QY      83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
Db      27010 GCTTTAAGTAATGCTCGAAGAAATTTCTGCAAGAAAGAAATGTAATAATTTACATGAAAAA 27069
QY      103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeu-----Glu 118
Db      27070 -----CAAATTCAGGAGCTCAACAAAGATGATGAG 27099

```

QY 119 AsPThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVal 138
|||||
Db 27100 GATACATAAGAT----- 27111
QY 139 LysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyr 158
::: |||
Db 27112 ---GCTTTAAAGCTATCCAAAGCTTTATCCAAAGCTAAAGCAATTCATACGTA 27168
QY 159 ArgThrLeuPheSerValIleAsnAsnGly-----HisAsnGlnLysValHisMetHis 176
::: |||
Db 27169 CAATCTTTA---TCTTATATCAATGATGCACAGCGCTCATATTGCTGTAATAATATATTCAC 27225
QY 177 MetValAspLeuMetSerSerIle-----IleCysGluGlyAspThr 190
||| ::|||
Db 27226 AACTCTGATGATTTATCATCTTTAGCAATATACATTATCTAAGCTAGATTTGATATAT 27285
QY 191 ValSerGlnIleLeuLeuAspThrVal-----LeuVal 201
::: |||
Db 27286 GCAATGAAAGCTTACGAGATACATAGAAAGTAATTCACCTTCTGTTCCAAATATAGTGT 27345
QY 202 AsnLeuValProAlaHisLysAsnLeuAsn-----LysGlnAlaTyrAspLeuAla 218
||| ::: |||
Db 27346 AATTATTTTATATGCTGATTAAGAAATTTTCAAAATTTGATGAGCGCTACAAAGCA 27405
QY 219 LysAlaLeuLeuLysArgThr----- 225
|||
Db 27406 AGTGCACAAGTCTTAAAGCTTCCAGAAATCCACAGATTCGAAGAATATAGTGT 27465
QY 226 AlGlnAlaIle-----GluProTyrIleThrThrPhe 236
::: |||
Db 27466 AGTCAGACCATTTACGATACAAAAATATGATTAATAGTGCACAGCACTTCGCAAGT 27525
QY 237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPhe 256
::: |||
Db 27526 AAGAGCAAGATCTAAATATATATATAA---CGATTAAAGATTTAAATTAAGCAACACT 27582
QY 257 AspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276
::: |||
Db 27583 GAGATGTCACAAAC-----AAGTAAATTCACCAAAATCTTTAAAGAGATTATCTTCG 27636
QY 277 Leu---GluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGln---ValValLys 294
||| ::: |||
Db 27637 CTCCTCAATCAACAGTTAGAAATTTAAACGATTAAGTAATATATGAGAGATTAAGCTTAA 27696
QY 295 LeuLeuAlaLysMetPheGlyAlaLys-----AspSerGluLeuAla 308
|||
Db 27697 ACTTTAGTAATCTGTTAAAGCAAGTTAAATATAGAAACGCTGATTTATATTTAAA 27756
QY 309 SerGln---AsnLysProLeuTyrGlnCys-----TyrLeuGlyArgPheAsnAspIle 325
||| |||
Db 27757 CGTCAATTTTAAACAAAGCTTTAAAGAGCTAAAGCGCTATTTAAATAAATAGCGGTACA 27816
QY 326 HisValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPro 345
::: |||
Db 27817 AATGTCATATTC-----AATGACATTCACAT---CTTTTAAACCAATA 27858
QY 346 AspLeuAlaLysAsp---LeuThrGluTyrLeuLysValArgSerHisAspProGln 364
||| |||
Db 27859 GATATGCTTAAAGCAATTAATGATGTAAGCAAGCTCTAAAGAACAATCAACAAATAATCT 27918
QY 365 AlaIle-----ArgHisAspValIleValSerIleValIlePheAla----- 377
::: |||
Db 27919 GAAGTATTTATTTTAAAGATTAAGATATCTTATATAGTCTCAAAAAGCTGCATATAT 27978
QY 378 -----AlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAspPheVal 393
||| |||
Db 27979 AATCAGATTTAGAGCGCTTAAAGACATTAATTAATTAATTAATTAATTAATTAATTAAT 28038
QY 394 ArgGluArgThrLeuAspLysArgTyrPargValArgLysGluAlaMetGlyLeuAla 413
28039 GAA-----TTAATATGCTATAGCAAGTTTA--- 28065
QY 414 GlnIleTyrLysLysTyrAlaLeuGln---SerAlaAlaGlyLysAspAlaAlaLysGln 432

Db 28066 -----AAGAACAATGTAGCTCATTTAAACACACACTACAAAGAC----- 28104
QY 433 IleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArg 452
::: |||
Db 28105 -----AACAATGAATTTTAAATGCTGATGAAGACCAT 28137
QY 453 LeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGln----- 469
||| |||
Db 28138 AAATTA-----CAATATGATTAACCTTCACAACTTAGCGAATATATGTT 28179
QY 470 -----ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
||| ::: |||
Db 28180 CTTGACAAGAAACGGTACAAATTAAGACGCTAATATCATATTTGAAATGATTCAAAC 28239
QY 484 LeuAsp-----LeuAsnAlaValLysAlaLeuAsnGlu---MetTrpLys 497
::: |||
Db 28240 ATGGATGATGCTAGACACTTCTTAAAGTAATGAATTAAGACTTAAAGATGCTCAAAACAAA 28299
QY 498 CysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLys 517
||| ::: |||
Db 28300 GCACATATATGCATTTAAAGATACGCTCAAAACGCTGATGAAATTTGAACACGCTAAT 28359
QY 518 ThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPro 537
::: |||
Db 28360 GCACATCAAAATCTTAAAGCT---CAAGCTTAAACAAATGTA----- 28398
QY 538 AspProGluLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspArgLys 557
|||
Db 28399 -----AATGAG 28404
QY 558 LysIleArgLysGlnLeuGluValLeuValSerProThrCysSer-----CysLysGln 575
::: |||
Db 28405 GAACCTGAAAGAGCGCTTTCTAATATATTAATGACCAACATCAATATGATTAGTAAACAA 28464
QY 576 Ala-----GluGly-----CysValArgGluIleThrLysLys 586
||| |||
Db 28465 GCAAAAGATGAAGGCAATCTGCAATTTGAACACATACATGCAGATTAATACCTAAAGCA 28524
QY 587 LeuLysAsnProLysGlnProThrAsnProPheLeuIleMetIleLysPheLeuGlu 606
::: |||
Db 28525 AAATCAGATGCTAATCAAAATGATGACCAAAAGCTTGAAGTATTAATCTTATATAGT 28584
QY 607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
::: |||
Db 28585 CAA-----AATCCAACTTATCAATGAGAAAGAAATAATTAATATGCTCAAAATTAAT 28638
QY 627 LysSerIleAspGlyThrAlaAspAspGluAspGluValProThrAspGlnAlaIle 646
||| ::: |||
Db 28639 AAGTTAGTAATGGAATTAAGAAATGAATTCACAAACGCTTAAACAAACAAATAGAA 28698
QY 647 ArgAlaGlyLeuGlnLeuLeuValLeuLeuSerPheThrHisAspProIleSerPheHisSer 666
||| |||
Db 28699 AATGCTCAACAAACATAGATGAAGCTATGAACTGTAAATAATTAATTAATGCGCAAA 28758
QY 667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
||| |||
Db 28759 GCAGAA----- 28764
QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
||| |||
Db 28765 ---GCTAAACAAATGATTAAGAAAGACTTATCCAAAGAAAGACAGAT----- 28806
QY 707 ArgSerAlaLeuLeuProValIleHisLysSerLysGlyProProAlaGlnAla 726
::: |||
Db 28807 -----GCAATTAATTAACACACTATTTTAACACCTTCTCAAAAGCA 28848
QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
||| ::: |||
Db 28849 -----CATGCTTTAGCAGATATGATTAACAGAAAGATGCACTTCACAT 28896
QY 747 IlePheGluProLeuHisLysSerLeuAspPro-----SerAsnLeuGlnHisLeuIle 764
||| |||

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/139,937
 FILING DATE: 20-OCT-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/979,156
 FILING DATE: 20-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: CAMPBELL, CATHRYN
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-CJ 9370
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-9001
 TELEFAX: 619-535-8949
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4868 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-139-937-12

Alignment Scores:
 Pred. No.: 2,52e-07 Length: 4868
 Score: 189.50 Matches: 277
 Percent Similarity: 35.50% Conservative: 247
 Best Local Similarity: 18.77% Mismatches: 608
 Query Match: 2.63% Indels: 345
 Gaps: 61

US-09-512-581b-2 (1-1391) x US-08-139-937-12 (1-4868)

QY 19 Vallysgluilleaseraplysile-----Serlysgluuglu 30
 Db 156 ATGAAGAATTTACACCAAACTCCATTACAGAGGTACAACTAATGACCAAAATTGAA 215
 QY 31 Metvalargargleuylsmevalvallysthphenetaspmetaspserglu 50
 Db 216 GCATGCATGAAATTTGGAATAATAGTTGGGGAACCTTAGAAGAAACCACTGATTAACT 275
 QY 51 Glugluylsgluuylr----- 56
 Db 276 GAAAAATTTGGAATATTTCTTGTGTGATCACACGAGCTTACTCCAGAGAGTAGAACTCTT 335
 QY 57 -----Leuasnleualaleuuhisleualaserasphepheleuylshisproglylys 74
 Db 336 GAAGGCTCATTTCTGATTAGAAATGCATGCAGAT-----AAATCATCAGCTGAA 386
 QY 75 Aspvalargleuleuvalalacyscysleuualaspllepheargtletyratalaproglu 94
 Db 387 GATATTGGAATTAATGTGGCC----- 407
 QY 95 Alaprotlythserproaspysleuylsaspillephenetheillethraaglinleu 114
 Db 408 -----AAGTGAATGACACCTGGAAGGAGAGAGATTCTTGTATGCGAAAAATGAGCTG 458
 QY 115 Lysglyleuglunspthryllysserproglinphenasnargtrrhetryrleuugluasn 134
 Db 459 AGTAGATCAGATCGAGAAAGCTAGCATTTGAGCATGAAAGCCCTTACCTGGAGGCTGAC 518
 QY 135 llealatrvallysserlyrnanillecysphegluileuglunasperasn-----glu 152
 Db 519 TTAGAGTAGTGAACAGAGAGATATGTTAGAAAAAGACAAATAAATAAGCAAG 578
 QY 153 llepethrghleuylrarghrhrlleupheservalilleasnlnghisasn----- 170
 Db 579 GTTATTGTCTGCTTGAAGAGAACTC-----TCAGTGTGTACAGTAGAGAGAAACCACTT 635
 QY 170 ----- 170

Db 636 CGTGGAATTTAGATCTATGTCAAAAAACCCAGGCACTGCATCAGTTCTGTAAAAA 695
 QY 171 -----gluysvalhismethismetvalaspleumetserseile--- 184
 Db 696 ATGAAGAGAAAAACAGAGAGCTGATGTCATCAAAAGAGAGTGTCCATTCATTCAG 755
 QY 185 llecysgluglyaspthryalsergin---gluleuleaspthryalvalleuvalasnleu 203
 Db 756 GTGGCAGAGCAGAGGTGAAGAAAGACCACTCTTGCAGCTTGTCTGTATGTC 815
 QY 204 valproalhhislysnleuasnlysglnalatyaspheulalaysalaleu----- 221
 Db 816 -----AGTAGCTGTTAAAAAGCAAAACATCTCCAGGAAAGCTGCAGACT 863
 QY 222 leuylsargthralaglnalaleugluprotyrletthrthphenasnghlvalleu 241
 Db 864 TTGGAAGAGACTCACAGAGACTGCTTTGACAAATGTGAGCTGGAACCAATTTGCA 923
 QY 242 Metleuglylys-----ThrserlaseraspseuSerghisval 255
 Db 924 CAACGTGAATTAAGAGAAAGATTTGTTGCAAGGAATCTGAAGCCCTGAGGCCAGACTG 983
 QY 256 pheaspheulleuugluuylrnanlleasprhishleuuleuSerValleupro 275
 Db 984 AGTGAATTCAGATTATGAAAAGCTGAATGTCTCCAGAGCCCTTGAGGCCGACTGTGAG 1043
 QY 276 gluleugluuphelsyleuylsSerasnaspnsgluu-----Argleu 290
 Db 1044 AAGGAGATGTCATTTAGAGCTGACCTCAACACAGAGAGAAAGTGCATCAGTACAGAA 1103
 QY 291 glinvallylsleuleualalylsmetpneghly-Alalylsaspsergluleualasergl 310
 Db 1104 GGCATGAGAACTGAGACTTCGATTTGAGGCCGATGAAGAAAGAGCTGCACATCGCA 1163
 QY 310 nasnlysproleutrglncystyrleugllyrargphenasnaspillehivalprollear 330
 Db 1164 GAGAAAC-----TGAAGAACCCGAGCGGGAATG-----ATTCACCTTAAGTAAAA 1211
 QY 330 glueuglucysvalylsphenelaserhiscys-----le 341
 Db 1212 GTTGAGAACCTTGAAAGGAAATTCAGAGATCTCGAAGAAACCCAGAGCTAGTATCTT 1271
 QY 341 umetasnhisproaspheulalaysaspheutrglucutyryleuylvalargserhissas 361
 Db 1272 GATG-----CCGAGATTTCCAAAGCAGAA-GTAGAGACTTAAACAAACAAATAGAGA 1324
 QY 361 progluuglualalearghissaspyalllevaliserillevalthalalayslysas 381
 Db 1325 GATGGCAGAGAGCTGAAGTTTGTGATTAGACCTTGTACAGCTTAAAGCTGAAAAAGA 1384
 QY 381 pilleuleuleuvalasn-----Asphisleule 390
 Db 1385 AAATCTGACAAACAATACAAAGAAACAGAGTCACTGTGTGAGACTGACAACTTACT 1444
 QY 390 uasnphenvalarggluargthrleuaspysatgtrp-----Argvalarglysgl 407
 Db 1445 CTCTTCATTAAAGTCTGTAGCAAGAAAGAGCAAGACAGACATCAATCAAAAGAA 1504
 QY 407 ualameetmegluleuaglnleuylrlys-----LysTyrAlaleuugliserlalel 425
 Db 1505 ATCTAAACATGCAGTGCAGATGCTTCAGATCACTTAAAGAGCTTAATGAGGCACTAC 1564
 QY 425 aglylsaspvalaleuylsghlalealatrpyrileysaspyrlyseuuhislyetrrly 445
 Db 1565 AGCCTTGTCGTGAGCAGAAATTTGAAGGCCACAGAAAGAGTGTGAGCCACCAAT 1624
 QY 445 rghnasnserlleaspsargleuleuvalgluatglllephelaglnltyrmetvalpr 465
 Db 1625 AGAGGAAGACATCAGCTGAGAAATAGCATTTGAAAGCTGAGAGCCGCG----- 1673
 QY 465 ohisasnleuugluThrThgluargmetlyscysleuylrtyrleuylalathleuas 485
 Db 1674 -----CTAGAGCTGATGAAGAGACAG----- 1697

QY	485	pLeuAsnAlaValLysAlaLeuAsnGluMetCtRlyScySclnSnLeuLeuAqHsGl	505
Db	1698	-CTGTGTCTTCAACAACAACACTGAAGAGAGAGCATGCATTCAGATTTTAAAGGGCTAG	1756
QY	505	nValLysAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa	522
Db	1757	AGTGGAGAACCTTGGAAAGAGAGCTTGAATGATGCCAGCAAAACCAAGACATGCAGCTCT	1818
QY	522	ILysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl	542
Db	1817	TGAGGACAGAAATTCCAAAGAGAGAGTGAAGCCCTAAAGCAAAAATAAGAAAGGATGAC	1876
QY	542	agLlnAspPhe-----MetLysLysPh	549
Db	1877	CCAAAGTCTGAGAGCTCTGAATTAGATGTTGTTACTATAAGTCCAGAAAAAATCT	1936
QY	549	eThrGlnValLeuGlnAspAspGluLysIleArgLysGlnLeuGlnValLeuAlaSerPr	569
Db	1937	GACAAATGCAATTACAAAAAGAGCAAGACGCAATCTGCAATAGAAATAATAATTCA--	1994
QY	569	oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs	589
Db	1995	-----TCATTTTGAAATAATTTTTCGACAAAAGAGCAAGCAAGAAAGTACA	2038
QY	589	nProLysGlnProThrAsnProPheLeuGlnMet-----IleLysPheLeu	605
Db	2039	GATGAAAGAAATAATCAACAGACTGCCATGCAATGCTTCAACACATTTAAAAGAGCTCA	2098
QY	605	uGlnArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLysGlnVa	625
Db	2099	TGAGAGAGATGGCAGCCCTGCATTAATGACCAAGACCGCTGAAGGCC-----AAAGACA	2152
QY	625	lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluLysIleProThrAspGlnAl	645
Db	2153	GAATCTTAGT-----AGTCAAGTAAGAGTCTCTGAACCTTGGAAGAGC	2194
QY	645	a---IleArgIlaGlyLeuGlnLeuLeuLysValLeuSerPheThrHisProIleSerPh	664
Db	2195	TCAGTGTCTACAAAGCCCTTGATGAGGCCAA-----ATAATTAATATTGTTT	2242
QY	664	eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLeuLysMe	679
Db	2243	GCAATCTTCAAGTAGAGCCCTCATTCACAAAGTAGAGATGAGCAAGCAACTGGAGCA	2302
QY	679	tAspAspGluLysValAlaGlnAlaIleLeuGlnIlePheLysAsnThrGlySerLysII	699
Db	2303	GAAGAGTGAAGAAATCAGTAGACAGCAAAATCAATT-----	2339
QY	699	eGluLysPheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis--	716
Db	2340	---CAAGACCAGAGCAGCTTGTCTCTAAACTGTCCAGGAGGAGAGAGCACCACACT	2395
QY	717	---LysSerLysGlyGlyProProArgGlnAlaLysThrAlaIleHisCysIleHisAl	735
Db	2396	TTGGAAAGAGCAAAACTTAGACTAGAGAAATCTGACAGTGAAATTGAGCAAGAAAGTCA	2455
QY	735	aIlePheSerSerLysGlnThrGlnPheAlaGlnIlePheGluProLeuHisLysSerIe	755
Db	2456	AGTCTCAATACCAAAATGCGCTTTCACAGACACATTAGAAAGTGGCAGAGTTCTTA	2515
QY	755	uAspProSerLysLeuGlnHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe	775
Db	2516	C-----AAGATCTTAGAGAT-----GAGCTTGAATT	2542
QY	775	lLeuAlaProAspGlnPheAlaAlaProTyrLysSerTrpValAlaThrPheIleVally	795
Db	2543	GACAAAAATGACAAATG-----TCCTTTGTT-----	2570
QY	795	sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs	815
Db	2571	-----GAAAAAGTAAACAAATGACTGCAAGGA	2599

QY	815	peGluslUvAlserProgluThrMetValLysllGlnAlAlleLysMetIcValArgTr	835
Db	2600	AACTGACGTCGACAGAGGAAATGCATGACATGGCACAGAAACACGACGACGTCGACAGAA	2659
QY	835	pleuLemlyMetLysAsn-----	845
Db	2660	ACTCGTGGAGAGAAATATAGCGTACGTCGAGAGTTCGACGTTACTGTTGGAGAAATATA	2719
QY	845	sSerGlyThrSerThrLeuArgLeuLeuThr-----	855
Db	2720	GAGCAGCAAAAGATCAATTGGAAGAGAGCTCACACTAGAAATAGTGAAATTGGAAGAGAGCTT	2779
QY	856	-ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe	875
Db	2780	AGATTGCATGCACAAAGAC---CAGGTGGAAAGAGAGGAAAGTG---AGAGAGCAAT	2833
QY	875	tSerArgLeuArgLeuAlAlaGlySerAlAlleValLysLeuAlaGlnGluProCysTy	895
Db	2834	AGCTGATATATAGCTACGACGGCTTCATGAGAGTGAAAGAA-----	2873
QY	895	rHisGluIleThrLeuGluGlnGlnTyGlnLeuCysAlaLeuAlAlaLeuAspGlyCys	915
Db	2874	-CACGAGCGCTTGCTTTGGAC-----	ACAAACAAACA 2905
QY	915	sTyGlnValArg--GlnValPheAlaGlnLysLeu-----HisLysGlyLeuSe	931
Db	2906	GTATGAAATGAATAATCCAGACATACCCGAGAGAAATGTACTTAAAGAAAGATGTCTCAG	2965
QY	931	rArgLeuArgLeuProLeuGluTyTyrMeAlAlleCys-----AlaLe	945
Db	2966	TTCACAGAGAGTGGAGATGACCTTTTAACTAGTAAAGAGAGCTCAATTAATCTATT	3025
QY	945	uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs	965
Db	3026	GAAGCTTACTACTCAGATTTTGGAGAAATTAGACAAACCAAGATGGCAATTTAAATA	3085
QY	965	nIleAsnValArgArgGluTyTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe	985
Db	3086	TGTTAAAT-----CAGTTGGAACAGAGAAATGACGTGCCACGGGGAATGAA	3133
QY	985	rLeuLeuProGluTyTyrValValProTyTThrIleHisLeuLeuAlaHisAspProAspTy	1005
Db	3134	GTTGTTG-----	3140
QY	1005	rValLysValGlnAspIleGluGlnLeuLysAspValLysGlnCysLeuTyTyrPheValLe	1025
Db	3141	-----ATCAAAATCTCTTAACACGCTGGAGAGGAAAGAGAGATCTGCAGAAAGAACT	3193
QY	1025	uGluIleLeuMetAlaLysAsnGluLysAsnSerHisAlaPheIleArg-----	1041
Db	3194	CTCTCAACTTCGACGTGCACAGAGAGAACAGAAACAGCTGTTATGATTACCAAGCT	3253
QY	1042	----LysMetValGlnAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys	1060
Db	3254	CGATGAATTAACAACCTGATGATCAAAAGAACTGGAAGAAACCTTGAGAGAAACCAAGGA	3313
QY	1060	sMetAsnGluLysLeuTyTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSe	1080
Db	3314	GGCAGATGAATACTTGGATAGTACTGT-----TCCTTGCTTATTAACCATCA	3361
QY	1080	rThrThrTySerLeuGlnUserProLysAsp-----ProValLeuProAlaArgPheP	1098
Db	3362	A-----AAGTTTACGAGAAAGCTAAAGAGATGTTAGAGACACACAGTCCCATCTGCTG	3412
QY	1098	eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyTyrLeuProProGluMetLysSe	1118
Db	3413	TTCACAGCAATCTAAACAAGATTTCCCGAGGCTCTCTTGCTAGTGCACAGTTGTT-----	3467
QY	1118	rPhePheThrProGluLysProLysTyThrThrAsnValLeuGluValAlaLysnLysProLe	1138
Db	3468	-----CCAGACACATCTCCAAATCCCTTCTGTTACTGAAAGAGAGCTT	3508
QY	1138	uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl	1158

```

Db      3509 ATCATCTGCCAAATTAACCTTCAGCAGAGGCAAGATCCAGTGAATATGCGAGAA 3568
QY      1158 aserSerserSersarProserProglYargIlelysgIlyArgleuaspserSergI 1178
Db      3569 TGGTGGAGGACCAACACCTGCTACCCAGAGACCTTTTCTAAAAAGCAAGAAAGCAGT 3628
QY      1178 uMetasp-----HissergluasngluaspTyrThrMetserSerProleuProgl 1195
Db      3629 CATTGAGTGTATTCACCTCGCAGAGACAGGAAAGTACAGTGTAGCCAGAGGAGCT 3688
QY      1195 yLysLysSeraspLysArgaspSerseraspLeuValArgSergIleuGlulysProAr 1215
Db      3689 TCCAGAGAGTTGTAAAGAGAGGTTTCTGACATCCAGCAAGAAAGACTAGCCATATAT 3748
QY      1215 gglYArgLysLysThrProValThrIgluIglu-----GluLysIgluI 1230
Db      3749 CTTGCGAAGAACACCAATGCGACATGGAGCAGCGCCGCCCTGCTGCACACAAGTTACC 3808
QY      1230 yMetaspaspLeuThr-----LysLeuValIgluIgluIlnlyspPro--LysGI 1245
Db      3809 CCTATCCCACTGACTGTCCCAACAAATCTTGCGAGAGTCTCCAAACCAACAGCTGG 3868
QY      1245 ySergIlnArgSersArgLys--ArgGIyHsThrAlaSergIleuSerspIgluIlntr 1264
Db      3869 TGGCAGCAGATCACAAAGGTGAAGTTGCTCAGCGGAGCCAGTAGATTCAGCAGCACAT 3928
QY      1264 pProgluIgluLysArgleuLysgluaspIleleuGlunlaspIgluInasn----- 1282
Db      3929 CTTCCGGAACCCACCCAGCAAAATCCGTCAGATCAATATCTTCTTAGAGAAATCCGAC 3988
QY      1283 -SerProLysLysglY-----LysArgGIyArpProProLysProleuGI 1298
Db      3989 TGACAGCCCCAGAGAGGCGCTGAGAGGTCAAGCGCGCGGCACTGTCCAGCCCCAAGC 4048
QY      1298 yglY-glyThrProLysgluIgluProThrMetLysThrSerLys 1312
Db      4049 TGGACTGTGAGTCCAAAGGAGTGAAGACTGTAAAGTCCAGTAA 4092

RESULT 11
PCT-US93-11310-12
; Sequence 12. Application PC/TUS9311310
; GENERAL INFORMATION:
; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11310
; FILING DATE: 19-NOV-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: FP-CJ 9790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4868 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US93-11310-12

Alignment Scores:
Pred. No.: 2,52e-07 Length: 4868
Score: 189.50 Matches: 277
Percent Similarity: 35.50% Conservative: 247
Best Local Similarity: 18.77% Mismatches: 608
Query Match: 2.63% Indels: 345
DB: Gaps: 61

US-09-512-581b-2 (1-1391) x PCT-US93-11310-12 (1-4868)
QY      19 ValLysgluIleSersapLysIle-----SerLysgluIglu 30
Db      156 ATGAAAGAAATTAGACTCAAACTCCATTACAGAGGTACACATTAATGACCAAAATTGAA 215
QY      31 MetValArgArgleuLysMetValValLysThrPheMetaspMetaspSergIu 50
Db      216 GCATGCATAGCATTTGAAAAAATATGTTGGGCACTTAAGAAAGAAACTCAGATTTAACT 275
QY      51 GluIgluLysgluLeuTyr----- 56
Db      276 GAAAAATTGGAAATTTTCTTGATCACCAGAGTTACTCCAGAGTAGAAACTTCT 335
QY      57 -----LeuasnleuAlaIleuHisleuAlaSeraspPheleuLysHisProglYlys 74
Db      336 GAAGCCTCAATTCGATTGAAGAAATCATGCACAT-----AAATCATCAGCTGAA 386
QY      75 AspValArgleuLeuValAlaCysCysleuAlaSpllePheArgIleTyrAlaProglu 94
Db      387 GATATTGAGATTAATGTGCCC----- 407
QY      95 AlaProTyrThrSerProaspLysleuLysaspIlePheMetPheIleThrArgIleu 114
Db      408 -----AAGGTGAATGACAGCTGAGAGAGAGATTTCTTGATGTGAAGAAATGACCTG 458
QY      115 LysglYleuGlunspThrLysSerProgluPheasnArgTyrPheTyrleuLeuGlun 134
Db      459 AGTAGATTCAGATCGGAAAGCAATGAGCATGAGCATGACCCCTCACTCGAGGCTGAC 518
QY      135 IleAlaTrpValLysSerTyrAsnIleCysPhegluLeuLysPserAsn-----Glu 152
Db      519 TTAGAGTATGTTCAACAGAGAACTATGTTTACAAAAACATGAAATTAAGCAGAG 578
QY      153 IlePheThrGlnleuTyrArgThrleuPheSerValIleasnAsnIyHisAsn----- 170
Db      579 GTTATTGTGCTGCTTGAAGAAAGCTC--TCAGTGTCAAGTAGAGAAACCAAGCTT 635
QY      170 ----- 170
Db      636 CGTGGAGAAATTAGATCTATGCAAAAAAACACGCGCACTGCATGAGTTGTGTAAGAA 695
QY      171 -----GlnLysValHisMetHisMetValaspLeuMetSerSergIle--- 184
Db      696 ATGAGAGAGAAACACAAAGAGCTGAGTCTCATCAAGTAGAGTCTCCATTCGATTGAC 755
QY      185 IleCysgluIglYaspThrValSergIln--GluLeuLeuaspThrValleuValasnLeu 203
Db      756 GTGGCAGAGCAGAGGTGAAGAAAGACGCACTCTTCAACTTGTGCTCTGAGGTG 815
QY      204 ValProAlahHisLysasnleuAsnLysGlnAlaTyrAspleuAlaLysAlaleu----- 221
Db      816 -----AGTGAAGCTGTAAAGAAAGCAAAATCATCTCCAGGAAAGCTGCAGAGT 863
QY      222 leuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheasnGlnValleu 241
Db      864 TTGAAAAAGACATCACAGGCACTGTCTTTACAAAATGTAGAGTGGAAAAACCAATTTGCA 923
QY      242 MetleuIgluLys-----ThrSerIleSersaspLeuSergIuHisVal 255

```



```

Db      924 CAACGAAATGAAGAAAGAAATTCCTGTGCAGGAATCTGAAGAGCTGCAGGCCAGACTG 983
      |||
      |||
Qy      236 PheaspLeuLleuGluLeuYrAsnIleasSerHisLeuLeuSerValLeuPro 275
      :|||
      :|||
Db      984 AGTGAATGAGATTAATGAAGAGCTGATCTTCACAGGCTTCGAGGCCGCACTGGTGGAG 1043
      :|||
      :|||
Qy      276 GluLeuGluPheLeuLysSerAsnAspAsnGluLys-----ArgLeu 290
      :|||
      :|||
Db      1044 AAGAGTGATTCGCTGAGCTGAGCTGACACACAGAGGAAGTGCATCAGCTGAGTGAAGA 1103
      :|||
      :|||
Qy      291 GluValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerG1 310
      :|||
      :|||
Db      1104 GGCATCGAAGAACTGAGAGTTGGCATGTGAGGCCGATGAAGAAAGACAGCTGCACATCGCA 1163
      :|||
      :|||
Qy      310 nasLysProLeuTrrpGlnCysTyrLeuGlyAArgPheAsnAspIleHisValProIleAr 330
      :|||
      :|||
Db      1164 GAGAAAC-----TGAAAGAACCGCGAGCGGAGAAATG-----ATTCACTTAAGGTAAAA 1211
      :|||
      :|||
Qy      330 GluGluCysValLysPheAlaSerHisCys-----Le 341
      :|||
      :|||
Db      1212 GTTGGAACCTTGGAAGGAATGCGAGATGTCAGAGAAACCGAGAGCTAGTGTCTT 1271
      :|||
      :|||
Qy      341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
      :|||
      :|||
Db      1272 GATG-----CCGAGAAATTCACAAAGCAGAA-CTAGAGACTCTAATAACACAAATGAGAGA 1324
      :|||
      :|||
Qy      361 prGluGluAlaIleArgHisAspValIleValSerIleValIthrAlaLysLysAs 381
      :|||
      :|||
Db      1325 GATGCCAGAACCCCTGAAAGTTTGTGAATTAAGCTGTACGTTAAGCTTGAAAGAAAGA 1384
      :|||
      :|||
Qy      381 pIleLeuLeuValAsn-----AspHisLeuLe 390
      :|||
      :|||
Db      1385 AAATCTGCAAAACAAATACACAGAAACCAAGCTAGTTCAGAACTAGACAACTTACT 1444
      :|||
      :|||
Qy      390 uAsnPheValArgGluArgThrLeuAspLysArgTrr-----ArgValArgLysG1 407
      :|||
      :|||
Db      1445 CTCTCATTTAAAGTCTGTAGAGAAAGAGACAGACAGATACAGATCAAGATCAAGAGA 1504
      :|||
      :|||
Qy      407 uAlaMetMetGlyLeuAlaGlnIleTyrLys-----LysTyrAlaLeuGlnSerAlaI 425
      :|||
      :|||
Db      1505 ATCTTAAACCTCAGCGAGATGCTGCAGAAATCAGTTAAAGGAGCTAAATGAGCGCAGTGA 1564
      :|||
      :|||
Qy      425 acGlyLysAspAlaIalLysGlnIleAlaTrrpIleLysAspLysLeuLeuHisIleTyr 445
      :|||
      :|||
Db      1565 AGCCTGTGTGTGTCAGCAAGAAATTAATGAAGCCACACACAGACTCTAGACCCACAT 1624
      :|||
      :|||
Qy      445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
      :|||
      :|||
Db      1625 AGAGCAAGAGCATCAGCTGAGAAATATGCAATGAAAGCTGAGAGCGCCG----- 1673
      :|||
      :|||
Qy      465 oHisAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
      :|||
      :|||
Db      1674 -----CTAGAACCTGATGAAGAAAGACAG----- 1697
      :|||
      :|||
Qy      485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrrpLysCysGlnAsnLeuLeuArgHisG1 505
      :|||
      :|||
Db      1698 -CTGTGTCTTACACAACTGAGAGAAAGTGAAGATGATGCAAGTTTACTTAAAGGTAG 1756
      :|||
      :|||
Qy      505 nValLysAspLeu-----LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
      :|||
      :|||
Db      1757 AGTGAGAAACCTTGAAAGAGAGCTGAGATGAGCCAGCAAAACCAAGAGCATGCGAGCTCT 1816
      :|||
      :|||
Qy      522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
      :|||
      :|||
Db      1817 TGAGCAGAGAAATTCAAAGAGAGAGCTGAGAGCCTTAAAGCAAAATAGAGGATGAC 1876
      :|||
      :|||
Qy      542 agLAspPhe-----MetLysLysPhe 549
      :|||
      :|||
Db      1877 CCAAGAGTGTGAGAGCTGTGAATTAAGATGTTTACTTAAGCTAGAGAAAGAAATCT 1936
      :|||
      :|||
Qy      549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValSerPr 569
      :|||
      :|||

```

```

Db      1937 GACAAATGAATTAACAAAAAGACAGAGCAATATCTGAATTAAGAAATTAATATCA-- 1994
      :|||
      :|||
Qy      569 oThrCysSerCysLysGlnAlaGluLysValArgGluIleThrLysLysLeuGlyAs 589
      :|||
      :|||
Db      1995 -----TCATTTGAAAAATATTTTTCAGAGAAAAAGACAGAGAAAGTACA 2038
      :|||
      :|||
Qy      589 nProLysGlnProThrAsnProPheLeuGluMet-----IleLysPheLeu 605
      :|||
      :|||
Db      2039 GATGAAGAGAAATTAAGAGACTGCGATGAGAGTGTCTTAACACAAATTAAGAGCTCA 2098
      :|||
      :|||
Qy      605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
      :|||
      :|||
Db      2099 TGAGAGAGTGGCAGCCCTGCATTAATGACCAGAAACCGTGAAGGCC-----AAGAGACA 2152
      :|||
      :|||
Qy      625 lAsnLysSerIleAspGlyThrAlaAspAspGluLysGluValProThrAspGlnAl 645
      :|||
      :|||
Db      2153 GAATCTTGTGT-----AGTCAACTAGAGTGTCTTGAACCTTGAGAAAGGC 2194
      :|||
      :|||
Qy      645 a---lLeArgAlaGluLeuGluLeuLysValLeuSerPheThrHisProIleSerPh 664
      :|||
      :|||
Db      2195 TCAGTTGCTACAGAGCCTTGATGAGGCCAA-----AATATTTATTTGTTT 2242
      :|||
      :|||
Qy      664 eHisSerAla-----GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
      :|||
      :|||
Db      2243 GCAATCTTCAGTGAATGGCCTTCATTCAGAGAGTAGAAGATGAGCAGAAACCTGAGAA 2302
      :|||
      :|||
Qy      679 tAspAspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysI1 699
      :|||
      :|||
Db      2303 GAAGGATGAGAGAAATCAGTACAGCTGAAATTCAAAT----- 2339
      :|||
      :|||
Qy      699 eGluLysPhePheProHisIleArgSerAlaLeuLeuProVal-----LeuHisHis- 716
      :|||
      :|||
Db      2340 -----CAAGACCAAGAGCAGCTTGTCTTAAGTGTGCCAGGTGGAAGAGACACCAACT 2395
      :|||
      :|||
Qy      717 ---LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
      :|||
      :|||
Db      2396 TTGGAAGAGCAAAACCTTAGAAACTGAGAAATCTGCAGATGGAATGAGAGCAAGATCCA 2455
      :|||
      :|||
Qy      735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLysHisLysSerLe 755
      :|||
      :|||
Db      2456 AGTGCTACAAATCCAAAAATGCGCTTTCAGAGACACATTAAGAGTGTGCGAGAGTTCTTA 2515
      :|||
      :|||
Qy      755 uAspProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe 775
      :|||
      :|||
Db      2516 C-----AAGATCTAGAGAAAT-----GAGCTTGAAAT 2542
      :|||
      :|||
Qy      775 lLeuAlaProAspGlnPheAlaAlaProTrrpLysSerTrrpAlaIleThrPheIleVal 795
      :|||
      :|||
Db      2543 GACAAAAATGACAAAAATG-----TCCTTTGTT----- 2570
      :|||
      :|||
Qy      795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrrpValProAs 815
      :|||
      :|||
Db      2571 -----GAAAAAGTAAACAAAAAGACTGCAAAAGCA 2599
      :|||
      :|||
Qy      815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
      :|||
      :|||
Db      2600 AACTGAGCTGCAGAGAGAAATGACATGACATGACACAGAAACGACAGCTGCAGAAAGA 2659
      :|||
      :|||
Qy      835 pLeuLeuGlyMetLysAsn-----AsnHisSerLys 845
      :|||
      :|||
Db      2660 ACTCAGTGAGAGAAATTAAGCTGAGAGTTCGAGAGTTCAGTACTTGTGAAGAAATGAAA 2719
      :|||
      :|||
Qy      845 sSerGlyThrSerThrLeuAlaGlyLeuThr----- 855
      :|||
      :|||
Db      2720 GAGCAGCAAGATCAATTAAGAGAGCTGCACCTAGCAAAATAGTGAATTAAGAGAGAGCT 2779
      :|||
      :|||
Qy      856 -ThrIleuHisSerAspGlyAspLeuThrGluGlnLysIleSerLysProAspMe 875
      :|||
      :|||
Db      2780 AGATTGCATGCACAAAGAC-----CAGGTGAAAGAGAGAGAGAGG-----AAGAGAGAAAT 2833
      :|||
      :|||
Qy      875 tSerArgLeuArgLeuAlaIalGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
      :|||
      :|||
Db      2834 AGCTGAATATCAGCTACGCGCTTCATGAGAGCTGAAAGAA----- 2873
      :|||
      :|||

```



```

DB:      3          Gaps:      50
US-09-512-581b-2 (1-1391) x US-08-836-022A-10 (1-19307)

OY      206 AlaHisLysAsnLeuAsnLysGlnAlaIleuAspLeuAlaLysAlaLeuLeuLysArgThr 225
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      12049 GCTTCAGAAACACTGAAACGCGGTGGACAGAAATTCCTGCCAATTCGTGTAGTGAAGAGT 11990

OY      226 Ala-----GlnAlaIleuGluProTyrIleThrIlePhePheAsnGlnValLeuMetLeuGly 244
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11989 AACTGGCTAGATGATCAACCAACATCATCTACCTTTTATATACAGTCAACCAATTTGGAA 11930

OY      245 -----LysThrSerIleSerAspLeuSerGlu----- 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11929 CAGATGACACTACTGCCGCAAAACTGTGTAACCAACCACTACCACTCATCAGAGCCA 11870

OY      254 HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerVal 273
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11869 ACAGCAATTAAGCCAGTTAAATAATTTGTAAAGATGAAGTCACAGATTTGTACGCTCTT 11810

OY      274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11809 CAGCTCAAAATTTGAG-----CAATTAAATAATTCAGACTCA 11774

OY      294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysPro 313
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11773 CAACCTGAAGAAAGAAAGGACAGGG-----CCA 11747

OY      314 LeuTyrGlnCysTyrLeuGluYarGpheaAsnIleHisValProIleArgLeuGluCys 333
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11746 ATGTTT-----CTGATGACAGACTTT 11726

OY      334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11725 GTGGCTTTACTAATCAT-----TTTAACCAAC----- 11699

OY      354 TyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11698 -----ATCTTTGAT 11690

OY      374 IleValThrAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
      ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11689 GGTGTAGGCGCCAAAGAGAAAGAGTACAGCAATTTTGTACACTTACCAACCAATGCCG 11630

OY      394 ArgGluArgThrLeuAspLysArgTyrPargValArgLysGluAlaMetMetLysLeuAla 413
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11629 TATCGAGAGCAATAGTAGC-----ATCAG----- 11603

OY      414 GlnIleTyrLysTyrAlaLeuGlnSerAlaIleGlyLysAspAlaLysGlnIle 433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11603 ----- 11603

OY      434 AlaTyrIleLys-----AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
      ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11602 ACGTGCATCCACAGCATGCAAGAAAGCAAACTCTGTACCTTATCTAGTCTAGTAATAT 11543

OY      452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThr 471
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11542 GAAATAAATGAGAGAGACTCGGAAATTTACAGGCTCCAAAGTTCTTTGAAA----- 11489

OY      472 GluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAla 491
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11488 GAGCAACAATAATGCGCTTCAACTATCTGAGTACACTGTGAAGAGATGGCCCAAGAAAGCA 11429

OY      492 LeuAsnGluMet-----TyrLys 497
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      11428 CTTTCAGAAATATGCCAGAAATATCTGTCAAGATTTGAAGAGATTTGAGGGCAGCTGGAAG 11369

OY      498 -----CysGlnAsnLeuLeuArgHisGlnValLysAsp 508
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      11368 AAATCTTCTCCCACTGGTGGAAAGAGTCCCAAAAGCTAGAGAACAT----- 11321

OY      509 LeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528

```

Db 10336 ATTGAAAGTCCTTCACCTTAATTCAGAGAGCTCGCTTAATTCATTCAGCAACAGCTGGCA 10277
 QY 820 ProgluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMet 839
 Db 10276 GCTTATTCACCTGACAGAGTGGATGAGCTCAATGCTTCAGAA-----GCC 10229
 QY 840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHis 859
 Db 10228 CAGAAATCCCAATCGAGATTGACACATGACATGAGATTAGTAAGAAATGAGAAACAT 10169
 QY 860 SerAspLysArgLeuThrGlnGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
 Db 10168 AACGAGGGAAGCATGCCAACCAAGAGCTTCTTCACAAATTTGATGTCACAGAAAAA 10109
 QY 880 LeuAlaIleLysSerAlaIleValLysLeuAlaGlnLysProCysTyrHisGln----- 897
 Db 10108 TTACAGATGCTCCACAGAAATTTGCATTTATCCAAAAACAGCCAAATTTTGAACACAGT 10049
 QY 898 -----IleIleThrLeuGlnGlnTyrGlnLeuCysAlaLeu----- 909
 Db 10048 CTAGAGAAAGTAAAGATGATTTTATGATGAAGTCAGATTCGCTGCAATTCGGAACCC 9989
 QY 910 ---AlaIleAsnAspGlyCysTyrGlnValArgGlnValArgLeuAlaGlnLysLeuHisLys 928
 Db 9988 AAGAGTGTGAACAGAGAAATTAATTCAGTCAACAACTAGCTATGTGTGAACCTGTATAA 9929
 QY 929 GlyLeuSerArgLeuArg----- 934
 Db 9928 AGCCTGAGTGAAGTCAGTCTGAAGTGAATGGTGTATAAACCGAGCTCAAAATTTGTA 9869
 QY 935 -----LeuProLeu 937
 Db 9868 CAGAAAAAGCAGACAGAAAAATCCCAAGAGCTTGATGACAGAGTAAACAGCTTTGCAATTTG 9809
 QY 938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
 Db 9808 CATTAACATGAGTTGGGTGG-----AAGTTAACAGAGAAAGCAACAG 9764
 QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlnTyrLeuLysGlnHisAla 977
 Db 9763 TTGGAGAAATCTTGAGATTGCTCCCGTAAGATGAGAAAGAA----- 9722
 QY 978 AlaValSerGluLysLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
 Db 9721 -----ATGAATGTCTTAAACAGAAATGG----- 9701
 QY 998 LeuLeuAlaHisAspProAspTyrValLys-ValGlnAspIleGlnGlnLeuLysAspVal 1017
 Db 9700 CTGGCAGCACAACAGATACAGAAATGACAGAAAGATCAGACAGTGT----- 9658
 QY 1017 LysGlnCys-----LeuTrpPheValLeu-----GluIleLeuMetAlaLysAs 1032
 Db 9657 -AAGGATGCCAAGTAAATTTGGATTCTGGAAGTTGCCCTGGGGAAGAGCTACTCAAAAGAG 9599
 QY 1032 ngLAsnAsnSerHisIlePheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
 Db 9598 ATTGAGAAACAG-AAGGCTCACTGGAAGAGCTGTACAGAAATGAGAGAGCTTTGGAAT 9540
 QY 1052 palGlnGlyProAspAspAlaLysMetAsnGlnLysLeuTyrThrVal----- 1068
 Db 9539 GGTGTGGGCAAGAAAGAACCTTGTAAGAAATAACTGAGTCTTGTGAACAGTAACTG 9480
 QY 1069 -----CysAspValAlaMetAsnIleIleMetSer----- 1078
 Db 9479 GATAGCTGTCACTCCAGAGTAGAAGAAATGCTAAATCTTTTGTGGATATACAGAAACA 9420
 QY 1079 ----LysSerThrThrTyrSerLeuGlnUserProLysAspProValLeuProAla---Ar 1096
 Db 9419 CATGGAACCTTTGATCAGAAATAGAAACAAATCACAAGTGGATCTATTATGACAGATGA 9360
 QY 1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluMet 1116
 Db 9359 ACITTTAGATGAGTGTGAAAAAGAAAGAACACAAACAAAGAAAGACATTTCTTAAGCGTTT 9300

QY 1116 LysSerPhePheThrProGluLysProLysThrThrAsnValLeuGlnAlaValAsnLys 1136
 Db 9299 AAGAGCTGAATGAATGATACATGCCCCCAAGGTGGATCTCCACACAGTACACAGACAGAAA 9240
 QY 1136 sProLeuSerSerAlaGly-----LysGlnSerGlnThrLys 1148
 Db 9239 ATTGATGGCAAAACCCCGGTGACACCTCCAGAAAGTAGTAGAGCCCAAAATCTTGAGCT 9180
 QY 1148 sSerSerArgMetClnThrValSerAsnAlaSerSerSerAsnProSerProGln 1168
 Db 9179 CAACCGTGAATTTGACACCTATTCTTCACAGAAATTAAGACGTGGAAAGCCCTCATTTCT 9122
 QY 1168 YArgGlnLysGlnArgLeuAspSerGlnMetAsnHisSerGlnAsnGlnLysArgTyrThr 1188
 Db 9121 -----TGAAGGAATTTGACACAGTTTAATCTCAGATTATCAAAA 9084
 QY 1188 MetSerSerProLeuProGluLys----- 1196
 Db 9083 ATTGCTTGAACCACTGAGGCTGAAATTCAGACAGCGGGTGAATTCGAAGAGAACATTT 9024
 QY 1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGlnLeuGlnLysProArgGln 1216
 Db 9023 CAATAAAGATATGATGATGACAGACATGAGGCTACTGTAAATGATTTGTCGAAGAGAGA 8964
 QY 1216 YArgGlnLysThrProValThrGlnGlnGlnGlnLysLeuGlnMetAspAspLeuThrLys 1236
 Db 8963 CAACCTTACAAACAAAGATACACAGATGAGAAAGCCGAGAGAAATAAAGATAAACAGCA 8904
 QY 1236 sLeuValGlnGlnGln-----LysProLysGlnSerGlnArgSerArgLysArg 1252
 Db 8903 GCTGTACAGACAAACAAATGAATGCTCAAGAGATTGAGGTCTCAAGAAAGAAAG-- 8846
 QY 1252 gGlnHisThrAlaSerGlnSerAspGlnGlnGlnTyrProGlnGlnLysArgLeuLysGln 1272
 Db 8845 -----GCCCTAAGAAATTTCTCACACCTGTATACAGTACAGACAGAGCTGCA 8799
 QY 1272 uAspIleLeuGlnAsnGlnAspGlnGlnAsnSerProProLysGlnLysArgGlnArg 1292
 Db 8798 TGATCTCTGAATGATCTTGATGAAATTT----- 8771
 QY 1292 gProProLysProLeuGlnGlnGlnThrProLysGlnGlnProThrMetLysThrSerLys 1312
 Db 8771 ----- 8771
 QY 1312 sLysGlnSerLysLysSerGlnProProAlaProGlnGlnGln-----GlnGln 1329
 Db 8770 -----GAAAAAAATTTACCAACCCCTACCTGAACCCAGAGATGAAGAAATTTAAAGCA 8718
 QY 1329 uGlnGlnArgGlnSerGlnSerGlnAsnThrGlnLysSerLysSerGlnHisArgValSe 1349
 Db 8717 AATTGATCGTGAATTTG-----CAGAAAGAAAGAAAGAGAGCTGATCCAGTGCAG 8670
 QY 1349 rArgArgAlaGlnGlnArgAlaGlnUserProGlnUserSerValIleGlnSerThrGln 1368
 Db 8669 CAGGCAAGCTGAGGGCTTGTCTGAGATGGGCGCGCAGATGCGAGTGGAGCCAACTCAG 8612

RESULT 13
 US-09-427-048A-10/c
 ; Sequence 10. Application US/09427048A
 ; Patent No. 6203975
 ; GENERAL INFORMATION:
 ; APPLICANT: Trustees of the University of Pennsylvania
 ; Wilson, James M.
 ; Fisher, Krishna J.
 ; Chen, Shu-Jen
 ; Wetzman, Matthew
 ; TITLE OF INVENTION: Improved Adenovirus Virus and
 ; Methods of Use Thereof
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Howson and Howson
 ; STREET: Spring House Corporate Cntr, P O Box 457

CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/427,048A
FILING DATE: 21-Oct-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,022
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNP/PN.008PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 19307 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

Alignment Scores:

Pred. No.: 8.87e-06 Length: 19307
Score: 184.50 Matches: 241
Percent Similarity: 33.38% Conservative: 213
Best Local Similarity: 17.72% Mismatches: 496
Query Match: 2.56% Indels: 411
DB: 3 Gaps: 50

US-09-512-581b-2 (1-1391) x US-09-427-048A-10 (1-19307)

QY 206 AlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThr 225
DB 12049 GCTTCAGAACACGACGACGCGGACAGCAATTCGCCAATTGCTGAGTGAGAGCTT 11990
QY 226 Ala---GlnAlaIleGluProTyrIleThrPhePheAsnGlnValLeuMetLeuGly 244
DB 11989 AACGGGCTAGAGTTCACACACACATCATTCCTTTTAAATCAGCTCAACAAATTGGAA 11930
QY 245 -----LysThrSerIleSerAspLeuSerGlu--- 253
DB 11929 CAGATGACAACTACTGCCGAAGAACTGTGTGAAGAACCCAGTCAACCCCTTACAGGCCA 11870
QY 254 HisValPheAspLeuIleLeuGlnLeuTyrAsnIleAspSerHisLeuLeuLeuSerVal 273
DB 11869 ACAGCAATTTAAAGCCAGTTAAATTTGTAAAGATGAAGTCAACAGATTTGTCACGCTCTT 11810
QY 274 LeuProGlnLeuGlnPheLysLeuLysSerAsnAspAsnGlnGluArgLeuGlnValVal 293
DB 11809 CAGCCTCAATTGAG-----CAATTAAATAATTCAAGTCTTA 11774
QY 294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlnLeuAlaSerGlnAsnLysPro 313
DB 11773 CAAGTGAAGAAAGGACAGGAG-----CCA 11747
QY 314 LeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGlnLys 333
DB 11746 ATGTTT-----CTGGATGACAGACTTT 11726
QY 334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu 353
DB 111 -----

DB 11725 GTGGCCTTTACTAATCAT-----TTTAACCAAC----- 11699
QY 354 TyrLeuLysValArgSerHisAspProGlnGluAlaIleArgHisAspValIleValSer 373
DB 11698 -----ATCTTGTAT 11690
QY 374 IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
DB 11689 GGTGTAGAGCCCAAGAGAGAGCTTACAGACAAATTTTGTACACTTTTACCAACCAATGGCC 11630
QY 394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGlnAlaMetMetGlyLeuAla 413
DB 11629 TATCAGAGACAAATGATGAC-----ATCAGG----- 11603
QY 414 GlnIleTyrLysTyrAlaLeuGlnSerAlaIleGlyLysAspAlaAlaLysGlnIle 433
DB 11603 ----- 11603
QY 434 AlaTrpIleLys-----AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
DB 11602 ACGTGATTCAGACAGTCAAGAAAGCAAACTCTGTACCTTATCTTATGTTACTGAATAT 11543
QY 452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlnThrThr 471
DB 11542 GAATTAATGAGAGAGAGACTCGGGAATTTACAGGCTCGCAAAAGTTCTTTGAAA----- 11489
QY 472 GluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAla 491
DB 11488 GAGCAACAATAATGGCTTCACTATCTGATGATGAGTGAAGAGAGATGCCAAGAAAGCA 11429
QY 492 LeuAsnGlnMet-----TyrLys 497
DB 11428 CTTTCAGAAATATGCCAGAAATATCTGCAAGATTTGAAGAGATTGAGGGCCACTGAAG 11369
QY 498 -----CysGlnAsnLeuLeuArgHisGlnValLysAsp 508
DB 11368 AAACCTTCTCCCACTGTGTGGAAAGCTGCCAAAAGCTGAGAAACAT----- 11321
QY 509 LeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
DB 11320 -----ATGAATTAACCTTCGAAAATTTTCAGAAATCAATTAACAAACCTTA---CAGAAA 11273
QY 529 ValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLys 548
DB 11272 TGGAG-----GCTGAAGTTGATGTTTCTCGTGAAGAG 11240
QY 549 PheThrGlnValLeuGluAspAspGlnLysIleArgLysGlnLeuGlnValLeuValSer 568
DB 11239 GAATGGCCTGCCCTGGGGATGCTGAATTCCTGAAAAACAGCTCAAA----- 11192
QY 569 ProThrCysSerLysGlnAlaGlnLysValArgGluIleThrLysLysLeuGly 588
DB 11191 -----CAATGACAGCTTTAGTTGATGATATTCAAAACAAATTCACCCAGTTTAAT 11141
QY 589 AsnProLysGlnProThrAsnProPheLeuGlnMetIleLysPheLeuLeuGlnArgIle 608
DB 11140 AGT-----GTTAAAGAAAGCTGGGAGAGATTAAGAGTGAAGCTGAAGCTTGAAG 11093
QY 609 AlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLysGlnValAsnLysSer 628
DB 11092 TTTCGATCCACAGTGAAGACAGAA-----CTTAGAGACTTAAACACTGAG 11048
QY 629 IleAspGlyThrAlaAspAspGlnLysGlnValProThrAspGlnAlaIleArgAla 648
DB 11047 TGGGATCAATATGC-----CGCCAGGTCTACACAGAAAGAGAGCCCTTAAAGCA 10997
QY 649 GlyLeuGlu-----LeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
DB 10996 GGTTCGATTAACACCCGTAAGCTCCAAAAGATCTATCAGAGATGATGATGATGATGACA 10937
QY 665 HisSerAlaGlnThrPheGlnSerLeuLeuAlaCysLeuLysMetAspAspGlu----- 682
DB 10936 CAAGCTGAAGAAAGATATCTGAGAGAGATTTTGAATTAACCTCCAGATGAATTAACAG 10877

QY 683 -----LysValAlaGluAlaLeuGln----- 690
Db 10876 ACTGCTGTGAAGAAATGAGAGAGCGTAAGAAGAGGCACTACAAAAGAACTAAAGTCG 10817
QY 691 -----IlePheLysAsnThrGlySerLysIleGluGlnAspPheProHISLeuArgSer 708
Db 10816 AACTCCCTTACTGACACTGTAAATAGTAAATAGCTCCACGCTCCACCCCTGAGCACAAGAG 10757
QY 709 AlaLeuLeuProValLeuHisHisLysSerLys----- 719
Db 10756 GCCTTTAAAAAGAACTTGAAACTGTGACCAACCAACTCAATGCGCTGTCACCAAGCTG 10697
QY 720 LysGlyProProArgGlnAlaLysTyraIleHisCysIleHisAlaIlePheSerSer 739
Db 10696 AATGGAATAATGCAAAACTTTGGAAAGAGTTGGGCACTGTGGCATGTGGCATGATTGTCTATAT 10637
QY 740 LysGlu----- 741
Db 10636 TTAGAGAAAGCAAAACAAGTGGCTCATGAAAGTAATGAAACTTAAACCAATGGAATAAT 10577
QY 742 -----ThrGlnPheAlaGlnIlePheGluProLeu----- 751
Db 10576 GTTCCTGACAGACCTGAGAAATCAGTCAGAGTGTGTAATCTCTGAAATCTGATGATCAT 10517
QY 752 HisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrProLeuValIThrIleGly 771
Db 10516 CATTCAGAGAGGAAACCAAACTGATTCGTCTATGTCACAGACATCTTACAGATGAGAGA 10457
QY 772 HisIleLeuLeuLeuAlaProAspGln-----PheAlaAlaProTyrPlySerTyr 788
Db 10456 GTCATGATGATCACTGATCAATGAGGAGCTTGAGACGTTGATTAATCTCTTGAGGGAACATA 10397
QY 789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThr 808
Db 10396 CATGAGAGGCTGTGAGGAACAACAACTGCTTGAACAAGATACCAAGTCTGCCACAGAA 10337
QY 809 ThrLys-----LeuTyrValProAspGluGlnValSer 819
Db 10336 ATTGAAGAAGTCCTTGCACTTAATTCAGAGAGTCCGTTGAAATTCATTTGACAAAGCAAGTGGCA 10277
QY 820 ProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrPheLeuGlyMet 839
Db 10276 GCTTATATCACTGACAGAGTGATGAGCTCAATGCGCTCAGAA-----GCC 10229
QY 840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHis 859
Db 10228 CAGAAATTCATCAATCAGATTGTGACAAAGTCATGAGATAGATTAGAAAGAAATGAAACAT 10169
QY 860 SerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
Db 10168 AACCCAGGGAAGCATGCCCAACCAAGCGTTCTTCCAAATGTGATGTTCGACACAGAAAAA 10109
QY 880 LeuAlaIleGlySerAlaIleValLysLeuAlaGlnLeuProCysTyrHisGlu----- 897
Db 10108 TTACAAAGATGTCTCCATGAAATTTCCGATTTTCCAAAAACCAAGCAATTTTGAACAAAGCT 10049
QY 898 -----IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeu----- 909
Db 10048 CTAGAGAGAAAGTAAAGATGATTTTAGATGAAGTCAAGATGCAATTTGCCCTGCAATTTGAAACC 9999
QY 910 ----AlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928
Db 9988 AAGAGTGTGTAACAGAAAGTAATTCATGACAACTAAGTCAATGTGTGCAACTTGTANAA 9929
QY 929 GlnLeuSerArgLeuArg----- 934
Db 9928 AGCCTGAGTGAAGTCAGAGTCTGAAGTGAATGGTGTATTAACCGGACGTCAAATTTGTA 9869
QY 935 -----LeuProLeu 937
Db 9868 CAGAAAAAGCAGACAGAAATCCCAAGAGCTTGATGACAGAGTAAACAGCTTTGAAATTTG 9809

QY 938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAlaHis 957
Db 9808 CATTTACATAGATGGGTGGCG-----AAGGTAAACAGAGAAAGCAACAG 9764
QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
Db 9763 TTGAGAAATAGCTTGAAGTGTCCCGTAGATGAGAAAGGA----- 9722
QY 978 AlaValSerGluLysLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
Db 9721 -----ATGAATGCTTTAAAGAAATG----- 9701
QY 998 LeuLeuAlaHisAspProAspTyrValLys-ValGlnAspIleGluGlnLeuLysAspVal 1017
Db 9700 CTGCAGCAACAACATACAGAAATTAAGACAAAGATGACAGCTTG----- 9658
QY 1017 LysGluCys-----LeuTyrPheValLeu-----GluIleLeuMetAlaLysAs 1032
Db 9657 -AAGGAATGGCAAGTAATTTGGATTCTGAAAGTTGCGGGGGAAGGCTACTCAAAAGAG 9599
QY 1032 nGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
Db 9598 ATTGAGAAACAG-AAGGCTCAGTTGAAGAGTGTACAGATTAAGAGAGCTTTGAAAT 9540
QY 1052 PAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrVal----- 1068
Db 9539 GGTGTGGGCAAGAAAGAACCTTGTAAGAGATTAACCTGACTTCTTGAACAGTAATCTG 9480
QY 1069 -----CysAspValAlaMetAsnIleIleMetSer----- 1078
Db 9479 GATAGCTGTCACTCCAGAGTAGAAAGATGGCTTAATCTTTGTGGAATACCAAGAAACA 9420
QY 1079 ----LysSerThrThrLysSerLeuGlnLysSerProLysAspProValLeuProLa--Ar 1096
Db 9419 CATGGAACCTTTGATGAGAAACAATAGAACAAATACAAACAAAGTGTATCATGCAAGATGA 9360
QY 1096 GphePheThrGlnProAspLysAsnPheserAsnThrLysAsnTyrLeuProGluMet 1116
Db 9359 ACTTTTGAAGAGCTGCAAAAAAGAAACCAACCAACAAAGAGACATCTTGAACGCTTT 9300
QY 1116 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuLysAlaValAsnLys 1136
Db 9299 AAGGCTGAATGANTACATGCGCCCAAGGTGAGACTCCACACGTCACAGACAGCAAA 9240
QY 1136 sProLeuSerSerLacly-----LysGlnSerGlnThrLys 1148
Db 9239 ATTGATGCAAAACCGCGGTGACCACTGACAGAAAGTAGTAGAGGCCAAATCTGTGAGCT 9180
QY 1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGln 1168
Db 9179 CAACCGTCGATTTGACAGCTATTTCTCCACAAATTAAGACAGTGAAGAGCCCTCCATTCCT-- 9122
QY 1168 YArgIleLysGlyArgLeuAspSerSerGluMetAsnHisSerGluAsnGlnLysArgTyrTh 1188
Db 9121 -----TTGAAGAAATTTGAGACAGTTTAACTCAGATATACAAA 9084
QY 1188 rMetSerSerProLeuProGlyLys----- 1196
Db 9083 ATTGCTTGAACCACTGAGGCTGAAATTCACAGAGGGGCTGAATCTGAAAGAGAGAGCTT 9024
QY 1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArgGln 1216
Db 9023 CAATTAAGAGATTAAGTGAACAAATAGAGAGTACTGTAAATGAATTTGTTGCCAAAGAGGAGA 8964
QY 1216 YArgGlyLysThrProValIThrGluGlnGlnGluLysLeuGlnLysMetAspAspLeuThrLys 1236
Db 8963 CAACCTTCAACAAGAAATCACAGATGAGAGAAAGCCGAGAGAAATAAAGTAAACAGCA 8904
QY 1236 sLeuValGlnGluGln-----LysProLysGlySerGlnArgSerArgLysAsp 1252
Db 8903 GCTGTTACAGACAAACATATATGCTCTCAAGAGATTTGAGGCTTCAAAGAAAGAAAAAG-- 8846
QY 1252 gGlyHisThrAlaSerGluSerAspGluGlnGlnIleTyrProGluGlnLysArgLeuLysGln 1272

Db 1414 -----CAGAGCTCCATCTCCAACTCAGCCAGGCCAAGAGAGCTGGAGCAGGCTCC 1467
 QY 254 -----HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu----- 269
 Db 1468 CAGGCTATGGGGCCGGTTGACATCGCCAGGTGGCCCTCTGACCTTGAAGCTCAGACACA 1527
 QY 270 LeuLeuSerValLeuProGlnLeuGluPheLys-----LeuLysSerAsnAspAsn 286
 Db 1528 CTCAAATGCTCCACATCGCAAGACAGATCAAGAACTGGCTGGCTGAAGCAGCAGGCCAAA 1587
 QY 287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
 Db 1588 GAGAGAGAGGCCAGCTGATGACAGACACCTC-----CAACAGCAAGAA 1629
 QY 307 LeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGluArgPheAsnAspIleHis 326
 Db 1630 CAGGCTCCAGGCGC----- 1644
 QY 327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
 Db 1645 -----CTCCGCCACACGGTGGAGCAGCTAAGCAGTGCCTGAAGCAGAAAGAGCAGCAG 1698
 QY 347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
 Db 1699 TTG---AAGAGGTAGCGGAG-----AAGCAGAGGCAACT 1731
 QY 367 ArgHisAspValIleValSerIleValThr----- 376
 Db 1732 AGGAGAGACCATGGCCAGCAACTGGCCACTGTCTGACAGAGCAGAGGCGCTCTTAAG 1791
 QY 377 ----- 380
 Db 1792 GAGCGGATCGCGCTCTCAAGCAGCTGAGCAGTGAAGAGAAAGAGGCTGCCAAGCTG 1851
 QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 Db 1852 GAGATTGTGACAGCAGCACTTCAAGTGTGATATGAGCCGGACAGCTGCCAGACCTCA 1911
 QY 401 ArgThrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr-- 419
 Db 1912 GTGACACAGGCCCGCGGAGAG-----GCAGAGCTGAGCCGGAAGGTGGAG 1959
 QY 420 AlaLeuGlnSerAlaAla-----GlyLysAspAlaAlaLysGlnIle 433
 Db 1960 GAATCTCCAGCGCTGTGTGAGACAGCCCGCAGAGAACAGCATGAGGCCAGGCGCAGGTT 2019
 QY 434 AlaThrIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspArgLeu 453
 Db 2020 GCAGAGCTAGAGTTGCGAGCTGCGTGTGAGCAGCAAAAGCACTGAGAAAGAAAGGTG 2079
 QY 454 LeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArg 473
 Db 2080 GCCCAGAGAGAAAGCAGCAGCTCCAGAGACAGTCCAGGCCCTCAAGAGTCTTAAGTGC 2139
 QY 474 MetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp-----LeuAsnAlaValLys 490
 Db 2140 ACCAAG-----GCCAGCCTTGAAGAGAGAGAGAGAGCCAGCGCTGCAGAT 2181
 QY 491 AlaLeuAsnGluMetTyrLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 510
 Db 2182 GCCCTGGAAGAGCAGCAGCGTTGTATCTGTGAGCTGAAGCAGAGACCCGGAAGCTGTGTG 2241
 QY 511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
 Db 2242 GAGCAGCATTAAGCGGAAGAGAGAG----- 2268
 QY 531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550
 Db 2268 ----- 2268
 QY 551 GluValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
 Db 2269 -----CTGAGAGAGAGAGGCGCTGGCGGCAAGGCGCTGGAGGCTCGATTACTCAGACTT 2322

QY 571 CysSerCysLysGlnAlaGlnGluCysValArgGluIleThrLysLysLeuGluAsnPro 590
 Db 2223 GGGAGAGCCCATCAGAGCTGAG-----ACTGAAGTCCCTGGCG----- 2358
 QY 591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluAlaGluIleAlaPro 610
 Db 2359 -----CGGAGCTGGCAGAGGCGCATGGCTGCC 2385
 QY 611 ValHisIleAspThrGlnUserIleSerAlaLeuLeuGlnValAsnLysSerIleAsp 630
 Db 2386 CAGCAGCAGCTGAGAGTGAAGTGTGACAGAGCTGTCTCAAGAGATAGCTGCTGGCGTGAAC 2445
 QY 631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 Db 2446 GGGTATGAGATACAGCAAGAGAGAGCAGATGATGGCCATGTTCCAGAGACAGCTG 2505
 QY 651 GluLeuLeuLysValIleSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
 Db 2506 ATGACTTTGAG-----GAGGAATGT 2526
 QY 671 GluSerIleLeuAlaCysLeuLysMetAspAspGluLysValAla-----Glu 686
 Db 2527 GAGAGGCCCGCCAGGAGCTGCAGAGAGCAAAAGAGAGTGGCAGGCAATGAATCCAC 2586
 QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheTronHisIle 706
 Db 2587 AGCAGAGTCCAGATTAAGCCCGGACAG-----AACAAACTAGTGTAGCTTCACCAACTG 2643
 QY 707 ArgSerAlaLeuLeuProValLeuHisIleLysSerLysGlyProValArgGlnAla 726
 Db 2644 GCCAGAGCACTCCAGCAGATC----- 2664
 QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
 Db 2665 -----CAAGAGAGAGATCAGGCGCCAGAAAG 2691
 QY 747 IlePheGluProLeuHis-----LysSerLeuAspProSer 758
 Db 2692 CTTCAGATGACCTCTGCACCTGTGACAGAAAGATGGCTGCACAGCAAAAGAGTGGCC 2751
 QY 759 AsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuLeuAlaPro 778
 Db 2752 CGCTTGAGAGACTTGGTC----- 2769
 QY 779 AspGlnPheAlaAlaProTyrLysSerTyrValAlaThrPheIleValLysAsp---Leu 797
 Db 2770 ---CGCAAGCGCAGGTGAGCAGCAGAGAAACAGCTCCGGGAGTTAGTCMAAGAGACTGCG 2826
 QY 798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrValProAspGluGlu 817
 Db 2827 AGGCGAGAGACAGACAGCCGAG-----TGGCTG-----GAAGAG 2862
 QY 818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyrLeuLeu 837
 Db 2863 CAACAGGAGAGCCAGTTCTGCAGACACAGCCAGCG-----CTGCAG 2904
 QY 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
 Db 2905 GCTATGAGCGGAGGAGCAGACAGACATGGCAATAGCTGGAACGGCTGCGGCGCGCTG 2964
 QY 858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
 Db 2965 ATGAGAGCCAGGCGGAGCAGACAGAGAGAGCGTGGCAGCAGAAAGAGAGTGGCGCG 3024
 QY 878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
 Db 3025 CTGACCCAGAGCGGCGCGCTGCCAGCTGACCTGACCTGCGTGGAG----- 3069
 QY 898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
 Db 3069 ----- 3069


```

Qy 918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
Db 3069 -----
Qy 938 GlnTyrMetAlaIleCysAlaLeuCysAlaIuAspProValLysGlnArgAlaHis 957
Db 3070 -----AAGCGCGCCAGACAGAG 3087
Qy 958 AlaArgGlnCysLeuValLysAsnIleAsnValAlaArgArgGlnTyrLeuLysGlnHisAla 977
Db 3088 CTGAGATGCGGCTGACAGACGCCCTCAACAGACAGCGCT-----GTGGAGTTCCT 3138
Qy 978 AlaValSerGlnLysLeuLeuSerLeuLeuProGlnTyrValValProTyrThrIleHis 997
Db 3139 ACCCTGCAAGAGGACATGCTCATGCTGACGGAAAAAGAA----- 3180
Qy 998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlu-----GlnLeuLys 1015
Db 3181 -----GGCAAGAGCACAGAGATTGGCCAAAGCTTGGTCTGAGAGCACCCAGATTAAAA 3234
Qy 1016 AspValLysGlnCysLeuTyrPheVal-----LeuGlnIleLeuMetAlaLysAsnGln 1033
Db 3235 GAGCTGGAGGAACCTTCGGCAAAACCTGAAGCACTGAAGAAACACACTGGCTTAAGAAAGAA 3294
Qy 1034 AsnAsnSerHisAlaPheIleArgLysMetValGlnAsnIleLysGlnThrLysAspAla 1053
Db 3295 AAGAGACACGATCTGGCTGACAGAGCCCAATCTGAGCTGCTGGAGAGACAGAG----- 3348
Qy 1054 GlnGlnLysProAspAspAlaLysMetAsnGlnLysLeuTyrThrValCysAspValAlaMet 1073
Db 3349 -----CCAAACAGGCCCCAACGCTGAAAGCACTGGCGGCGAGAGTGAGCAACCTGAAACAG 3402
Qy 1074 AsnIleIleMetSerLysSerThrThrLysSerLeuLeuSerProLysAspProValLeu 1093
Db 3403 CAATGCCCAAGAACAGACAGAGAGCTGACAGCTGGAACCAAGCTTGCAGAGCTGAGCGG 3462
Qy 1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
Db 3463 GCTTCGCCGG-----GCTGAGCGGGAGACAGTCTCTGAGACTCTGACAGGCCAGTTAAG 3516
Qy 1114 ProGlnMetLysSerPhePheThrProGlnLysProLysThrThrAsnValLeuGlnVala 1133
Db 3517 GAGAGAGGCCAGAGAG-----CTAGGGCGAC 3540
Qy 1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGln 1153
Db 3541 AGTCAGAGTGCCTTAGCTTCGCCCCAACGGGAGTTGCTGCTCCGACCAAGGTATAC 3600
Qy 1153 uThrValSerAsnAlaSerSerSerAsnProSerSerProGly-----ArgGln 1170
Db 3601 GACCAACAGACAGAGCTGAAGATGAGTGAAGGCCCAAGTGGCCCGGCGGCAAGAGCT 3660
Qy 1170 eLysGlyLysArgLeuAspSerSerGlnMetAsp-----HisSerGlnAsnGlnVala 1186
Db 3661 GAGAGGAAAAATATACCTCATCATGAGCACTTGAGAGAGAGAGAGTGCATCTCTGAAA----- 3713
Qy 1186 spTyrIleMetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAsp 1205
Db 3714 -----TCCGACAGTCTCTGAGAAAGAGGGGAGAGCAAGAGTTCGAAGCGG 3759
Qy 1206 LeuValArgSerGlnLeuGlnLysProArgGly-----Arg 1217
Db 3760 CTGGTGAATGGCCGAGTCAAGAAAGACCAAGAGCTGAGAGAGAGCTGGCTGCTGACAGG 3819
Qy 1218 LysLysThrProValIleThrGlnGlnGlnGlnLysLeuGlnMetAspAspLeuThrLysLeu 1237
Db 3820 CAGAGACAGCCAGCAACAGTCCAGAGCTGCAG-----AACGACGCTCG 3864
Qy 1238 ValGlnGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSer 1257
Db 3865 CTCTGCGGAGAGAGTGCAGAGCTTCGCGGAGGAGAGCTTGAAACACGCGGCTTCA 3924
Qy 1258 GluSerAspGlnGlnIntPrProGlnLysArgLeuLysGlnAspIleLeuGlnLysn 1277

```

```

Db 3925 GAGAACTCGCGCAGAGACTGACCTCAAGAGCTGAGCTGGCGAGAGACTG-----GGC 3978
Qy 1278 GluAspGlnGlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLe 1297
Db 3979 CAAGAAATTGAAGGCGTGCAGAGAGAGTTCTTCAGAAAGAGACAGGCCCTCTCCACCTG 4038
Qy 1297 uGlyGlyLysThrProLysGlnGluPro-----ThrMetLysThrSerLysLys 1313
Db 4039 CAGCTCGAGCACACACAGACAGAGCCCTGTGATGATGAGTGCCTGCGAGCTTAAGCACTC 4098
Qy 1313 sGlySerLysLysLysSerGlyProProAlaProGlnGlnGlnGlnGlnGlnGlnGln 1330
Db 4099 TGCCAGAGCTGCAGAGCCGAGAGAGCCCTGCTCCGAAAGAACCCACCTGAGAGAGCTGAG 4158
Qy 1330 ----- 1330
Db 4159 CAGAGACAGAGCGCGTGGGAGACTGCGGCGAGAGCTGCTGCGGCGCCAGCGGAGCTT 4218
Qy 1331 -----GluArgGlnSerClyAsnThrGlnGlnLysSerLysSe 1343
Db 4219 GGGAGCTGATTCCTTCGCGCAGAGAGTGGCAGAGACAGACCGAAGCTTCAGCAGCTG 4278
Qy 1343 LysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlnSerSera 1363
Db 4279 CGGGCAGAGAGAGCCAGCTATGACAGAGAGCTGAGCAGTGCAGAAAGCCGATGGCCTG 4338
Qy 1363 lAlleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
Db 4339 CT-----GGCAGAGGAGAAACCG 4355

RESULT 15
PCT-US93-06160-3
; Sequence 3, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: TESTA HURWITZ & THIBEAULT
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:

```

```

;
; AUTHORS: COMPTON, DUANE A
;
; AUTHORS: SZILAK, ILIYA
;
; AUTHORS: CLEVELAND, DON W
;
; TITLE: PRIMARY STRUCTURE OF OF NIMA, AN INTRANUCLEAR
;
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
;
; JOURNAL: JOURNAL OF CELL BIOLOGY
;
; VOLUME: 116
;
; PAGES: 1395-1408
;
; DATE: MAR-1992
;
; PCT-US93-06160-3

Alignment Scores:
Pred. No.:          1.9e-06          Length:          6306
Score:              182.50          Matches:          281
Percent Similarity: 31.34%          Conservative:    195
Best Local Similarity: 18.50%       Mismatches:      549
Query Match:        2.54%          Indels:          494
DB:                  Gaps:          55

US-09-512-581B-2 (1-1391) x PCT-US93-06160-3 (1-6306)

QY      16  PropGluVallylsGluIleSerAspLysIleSerLysGluGluMetValArgArgLeu 35
      |||  :::::|||||:::  |||||  |||  ::|  |||||
Db      823  CCCAAGAGCGCTGAGAGCTGCGTGCACAAG-----AATGAGAGCCTTACCATGCGGCTG 876

QY      36  LysMetVallylsThrPheMetAspMet----- 45
      |||  ::|||
Db      877  CATGAAACCTGTGAAGCGCTCCAGACCTTACAGACAGAGAAAGACCGCATGATCGCAA 936

QY      46  ---AspGlnAspSerGluGluGluLysGluLeuTyrlleuAsnleu-----Alaleu 61
      ::|||  |||||  |||  ::|||  |||
Db      937  ATCAACACAGCTTTCGGAGAGAAATGAGACACTTCTTAAGCTGGCGGAGTTGGCAGT 996

QY      62  HisLeuAlaSerSerPhePheLeuLysHisProGluLysAspValArgLeuValAla 81
      |||||
Db      997  CATCTG----- 1002

QY      82  CysCysLeuAlaAspIlePheArgIleTyrlAlaProGluAlaProTyrlThrSerProAsp 101
      |||||
Db      1003 -----CAG 1005

QY      102  LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLysAspThrLys 121
      ::|||  ::|||  |||  ::|||
Db      1006  CACACTCAGAGTGCCTCAATGAGCTGACGCGAG-----GAGCACAGCAAG 1050

QY      122  SerProGlu-----PheAsnArgTyrlPheTyrlleuGluAsnIleAlaIleTyrlLys 139
      ::|  |||  ::|  |||  ::|  |||
Db      1051  GCCACTCAGAGTGGCTAGAGAGACAGCCAGCTGGAGAGAGCTCAGCGCAGCCCTG 1110

QY      140  SerTyrlAsnIleCysPheGluLeuGluAspSerAsnGluIle-----PheThr 155
      |||  |||||  |||  |||||  |||  |||||
Db      1111  CAGGACACAGAAATGC-----CTTGAAGAGAGAGCAAAATCCTTCAGGAAAACTTCA 1164

QY      156  GlnLeuTyrlArgThrIleuPheSerValIleAsnAsnGlnLysAsnGlnLysValHisMet 175
      |||||  |||  ::|  |||  ::|||  |||  ::|||
Db      1165  CACCTGGAGAACACTGTGCTCCAGCTGCAGATAACCCACCCAGGAGGAGGCGAGGTG 1224

QY      176  HisMetValAspLeuMetSerIleIleCysGlnGluAspThrValSerGlnGluLeu 195
      |||||  |||||  |||||  |||||  |||||
Db      1225 -----CTGGGTGATGTCTTGGACGTGGAACC 1251

QY      196  LeuAspThrValIleuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrl 215
      |||  |||  ::|  |||  |||  |||
Db      1252  TTGAAGCAAGAGGACGCACTTTCGCGCAAAACACACACACAGCTCCAGGAGGTAGAG 1311

QY      216  Aspleu-----AlaLysAlaLeuLeuLysArg----- 224
      |||  |||  |||||  |||  ::|||
Db      1312  ATCTGTGAGACTGAGCGAGCGCAGCAAGAGCCAAAGCTGCTGTGCGCGGCGCACTTC 1371

QY      225  -----ThrIleGlnAlaIleGluProTyrlIleThrThrPheAsnGlnValLeuMet 242
      |||  ::|||  |||||  |||||
Db      1372  GAAGAAGAAAGACGAGCTGTCTAGCCTGATCATGACCTG----- 1413

```

```

QY      243  LeuGluTyrlThrSerIleSerAspLeuSerGlu----- 253
      ::|||  |||||  |||||  |||||  |||||
Db      1414  ---CAGAGCTTCATCTCAACTTACCTGAGCAGCCAGGACAGGAGAGCTGAGCAGGCTCC 1467

QY      254  -----HisValPheAspIleuIleLeuGluLeuTyrlAsnIleAspSerHisLeu----- 269
      |||  |||  ::|||  |||  ::|||  |||  |||
Db      1468  CAGGCTCATGGGCGCGTGTGACTGCCCATGGTGGCGCTGTGACCTGTGACCTGTGACCTGAC 1527

QY      270  LeuLeuSerValLeuProGlnIleGluIlePheLys-----LeuLysSerAsnAspAsn 286
      |||  ::|  ::|  |||  ::|  |||  |||||
Db      1528  CTCATGTCACACCTCCAGCAAGAGATCAAGCAACTGGCTGCTGAAGCAGCAGGCCCAA 1587

QY      287  GlnGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGluAlaLysAspSerGlu 306
      |||  ::|||  |||||  |||||  |||||
Db      1588  GAGAGAGAGGCGCCAGCTAGCAGCAGACCTC-----CAACGACAGAA 1629

QY      307  LeuAlaSerGlnAsnLysProLeuTyrlGlnCysTyrlleuGluArgPheAsnAspIleHis 326
      |||||  |||||
Db      1630  CAGGCTCCAGGCG----- 1644

QY      327  ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
      ::|||  ::|  ::|  |||  ::|||
Db      1645  ---CTCCGCCACAGCTGAGGAGCAGCTAAGAGATGCGCTGAAGCAGAGAGCAGCAG 1698

QY      347  LeuAlaLysAspLeuThrGluTyrlleuLysValArgSerHisAspProGluGluAlaIle 366
      |||  |||||  |||  |||
Db      1699  TTG---AAGAGAGTGGGAG-----AAGCAGAGGCACT 1731

QY      367  ArgHisAspValIleValSerIleValThr----- 376
      |||  |||  ::|  |||
Db      1732  AGCGAGACCATGCGCCAGCAACTGCGCACTGTGCAGAGAGAGAGAGCGCTCTTAAG 1791

QY      377 -----AlaAlaLysLys 380
      |||||
Db      1792  GAGCGGATCGCGCTCTCAAGCAGCTGAGGACCTGAGAGAGAGAGAGAGCTGCGCAAGCTG 1851

QY      381  AspIleLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
      ::|||  |||||  |||  |||||
Db      1852  GAGATTCGACAGCAGCAACTTCAAGTGTGTAATGAAAGCCGGGAGAGTGGCCAGACCTCA 1911

QY      401  ArgTrpArgValArgLysGluAlaMetMetGluLeuAlaGlnIleTyrlLysTyrl--- 419
      ::|  ::|  ::|  |||  ::|||  |||  ::|||
Db      1912  GTGCACACAGCCCGCAGCGGAGAG-----GCAGAGCTGAGCCCGAAGGTGGAG 1959

QY      420  AlaLeuGlnSerIleAlaIle-----GlyLysAspAlaAlaLysGlnIle 433
      |||||  ::|||
Db      1960  GAACCTCAGGCTGTGTGAGACAGCCCGCAGAGACAGATGAGGCCAGGCCAGGTT 2019

QY      434  AlaTrpIleLysAspLysLeuLeuHisIleTyrlTyrlGlnAsnSerIleAspAspArgLeu 453
      |||  ::|||  ::|||  |||  ::|||  |||  ::|||
Db      2020  GCAGAGCTAGAGTTCGAGCTGCGGCTGAGCAGACAAAAACCACTGAGAAAGAGGTG 2079

QY      454  LeuValGluArgIlePheAlaGlnTyrlMetValProHisAsnLeuGluIleThrGluArg 473
      ::|  |||  |||  ::|  |||  |||  ::|||
Db      2080  GCCCAGGAGAGAGCCAGCTCCAGAGCACCTCCAGGCTCCCAAGATGCTCTTGAAGTTC 2139

QY      474  MetLysCysLeuTyrlTyrlleuTyrlAlaThrLeuAsp-----LeuAsnAlaValLys 490
      |||  |||||
Db      2140  ACCAAG-----GGCAGCCTTGAAGAGAGAGAGAGCCAGCGGTGCGAGAT 2181

QY      491  AlaLeuAsnGluMetTyrlLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeu 510
      |||||  |||  ::|||  |||  ::|||  |||  ::|||
Db      2182  GCCTGTGAAGAGAGCAGCGCTGTGTATCTGTGACTGAAGCAGAGACCCGAAAGCTGTG 2241

QY      511  AspleuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
      ::|  |||  ::|  |||  ::|  |||
Db      2242  GAGCAGCATTAAGCGGAGCAAAAGAG----- 2268

QY      531  ValIleThrArgAsnLeuProAspProGluLysAlaGlnAspPheMetLysLysPheThr 550
      |||  ::|||  |||||  |||||
Db      2268 ----- 2268

```

QY 551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
 Db 2269 -----CTGGAAGAGAGAGGGCTGGCGCAAGGGGCTGGAGCTCGATTACTGACGCT 2322
 QY 571 CysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro 590
 Db 2333 GGGGAGGCCCATCAGCTGAG-----ACTGAAGTCTCTGGC----- 2358
 QY 591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuGluArgIleAlaPro 610
 Db 2359 -----CGGAGCTGGCAGAGAGCCATGGCTGCC 2385
 QY 611 ValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
 Db 2386 CAGCACACAGCTGAGAGTGAAGTGAAGCAGCTCGCAAGAAGTAGTCCCTGGGGGTGAC 2445
 QY 631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
 Db 2446 GGGTATGAGATGACGACCAAGAGGACAGTATGGCCCATGTTCCAGACAGACGCTG 2505
 QY 651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
 Db 2506 ATGACTTTGAC-----GAGCAATGT 2526
 QY 671 GluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla-----Glu 686
 Db 2527 GAGAAAGCCCGCCAGAGCTGACAGGCAAGAGGAGAGAGGCTAGAACATCCAC 2586
 QY 687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
 Db 2587 AGCGGATCTCCAGATAGCCGGGACAG-----AACAACTAGCTGACTCCATCCCAACCTG 2643
 QY 707 ArgSerAlaLeuLeuProValLeuHisIleLysSerLysGlyProProArgGlnAla 726
 Db 2644 GCCAAGACACTCCACAGACTC----- 2664
 QY 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerLysGluThrGlnPheAlaGln 746
 Db 2665 -----CAAGAGAAAGATGACGGCCCAAG 2691
 QY 747 IlePheGluProLeuHis-----LysSerLeuAspProSer 758
 Db 2692 CTTCAGATGACCTTCACCTCTGCAGAAAGATGGCTGCCACGCAAGAGGTGGCC 2751
 QY 759 AsnLeuGluHisIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPro 778
 Db 2752 CGCTTGGAGACTTGCTC----- 2769
 QY 779 AspGlnPheAlaAlaProThrLysSerTyrValAlaThrPheIleValLysAsp---Leu 797
 Db 2770 ---CGCAAGGCGAGTGAAGAGAAACAGCCCTCCGGAGTTAGTCAAGAGCCTGGC 2826
 QY 798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrValProAspGluGlu 817
 Db 2827 AGGGCAGAGACAGACAGCCCGAG-----TGCGTCTG-----GAGAG 2862
 QY 818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeuLeu 837
 Db 2863 CAACAGGAGACGCTTGCAGCACACAGAGCGG-----CTGCAG 2904
 QY 838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
 Db 2905 GCTATGAGCGGAGGAGCAGAGACAGATGGGCAATGAGTGAAGCGCTCGGGCGCGCG 2964
 QY 858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
 Db 2965 ATGAGAGCCAGGAGGAGCAGACAGAGGAGCGCTGGCAGCAGAAAGAGTGGCGCG 3024
 QY 878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
 Db 3025 CTGACCCAGAGAGGGGGCGTGGCCAGGCTGACCTTGGCTGGAG----- 3069
 QY 898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917

Db 3069 ----- 3069
 QY 918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
 Db 3069 ----- 3069
 QY 938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
 Db 3070 -----AAGCGCGCCAGACAG 3087
 QY 958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
 Db 3088 CTGAGATGGCGCTCAGAACGCCCTCAACAGGACAGCGT-----GTGAGTTCGCT 3138
 QY 978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
 Db 3139 ACCCTGCAGAGAGCACTGGCTCATGCCCTGACGGAAGAA----- 3180
 QY 998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlu-----GlnLeuLys 1015
 Db 3181 -----GGCAAGACACAGAGAGTTGGCCAAAGCTTGGCTGTGGAGAGCCCAAGTATAAA 3234
 QY 1016 AspValLysGluCysLeuThrPheVal-----LeuGluIleLeuMetAlaLysAsnGlu 1033
 Db 3235 GAGCTGAGAGAACTTCGCCAAACCGTGAAGCAACTGAAGAGAAACACAGCTGCTAAGAAAGA 3294
 QY 1034 AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla 1053
 Db 3295 AAGGACAGCGATGTGGCTCAGAGCCCAATCTGAGGCTGTGGCAGGACAG----- 3348
 QY 1054 GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValLysAspValAlaMet 1073
 Db 3349 -----CCAACAGGCCCAAGCTGGAAGACACATCGCGGAGAGTGAAGACAG 3402
 QY 1074 AsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeu 1093
 Db 3403 CAATCCAGAACACAGACAGAGCAGCTGACACTGGAACCGACGCTGAGCAGCTGAGCGG 3462
 QY 1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
 Db 3463 GCCCTCCGG-----GCTGAGCGGAGCAGAGTGTCTGAGACTCTCAGGCGCAGTTAGAG 3516
 QY 1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAla 1133
 Db 3517 GAGAAAGGCCACAGAG-----CTAGGCGAC 3540
 QY 1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr---LysSerSerArgMetGln 1153
 Db 3541 AGTCAGAGTGCCTTAGCTCGGCGCAACGGAGTTGGCTGCTCCGCCCAAGGTACAA 3600
 QY 1153 ThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly-----ArgIle 1170
 Db 3601 GACCAACACCAAGGCTGAAGATGATGGAAGGCCAGTGGCCCGGGCGCGCAGACAGCT 3660
 QY 1170 eLysGlyArgLeuAspSerSer---GluMetAsp-----HisSerGluAsnGluA 1186
 Db 3661 GAGAGGAATAATAGCTCATTCAGCACTTGGAGGAGAGGTTCATCTGAA----- 3713
 QY 1186 sPtyrThrMetSerSerProLeuProGly---LysLysSerAspLysArgAspSerAsp 1205
 Db 3714 -----TCGCCAGGTCTCTGGAGAAAGAGGGGAGAGACAGAGAGTTGAACGG 3759
 QY 1206 LeuValArgSerGluLeuGluLysProArgGly-----Arg 1217
 Db 3760 CTGGTGATGGCGGATCAGAGAGCCAGAAAGCTGAAGAGAGAGCTGGCGCTGTGCAGG 3819
 QY 1218 LysLysThrProValThrGluGlnGluLysLysLeuGlyMetAspAspLeuThrLysLeu 1237
 Db 3820 CAGACACAGCCAGCAACAGTGCAGAGCTGCAG-----AACCACTCTG 3864
 QY 1238 ValGlnGluLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSer 1257

```
Db 3865 CTCTGCGGAGGAGGTGACAGCCTCCGGAGGGAGGCTGGAATAACAGCGGGTGGCTTCA 3924
QY 1258 GluSeraspIuIngnIInTPProgluGluLysArgLeuLysGluaspIleLeuGluAsn 1277
Db 3925 GAGAACCTGCGGCGAGCTGACCTCACAGGCTGAGCGCTGGGAGAGCTG-----GGC 3978
QY 1278 GluaspIuInasnSerProProLys-LysGlyLysArgGlyArgProProLysProLe 1297
Db 3979 CAGAATTGAAGGGGTGGCAGGAGAAATTCTCCAGAAAGACGAGGCCCTCTCCACCCTG 4038
QY 1297 uGlyGlyGlyThrProLysgluPro-----ThrMetLysThrSerLysLy 1313
Db 4039 CAGCTCGAGCACACACAGCAGGCCCTGTGAGTAGAGCTGCTGCCAGCTAAACACCTC 4098
QY 1313 sGlySerLysLysSerGlyProProAlaProgluGluGluGluGlu----- 1330
Db 4099 TGGCAGCAGCTGCAGGCCGAGCAGGCCGCTGCGAATAACGCCACCGTGAAGAGCTGGAG 4158
QY 1330 ----- 1330
Db 4159 CAGAGCAAGCAGGCCGCTGGGGGACCTGCGGGCAGAGCTGCTGGGGCCAGCGGAGCTT 4218
QY 1331 -----GluArgGlnSerGlyAsnThrGluGlnLysSerLysSe 1343
Db 4219 GGGGAGCTGATTCCTTCGCGCAGAAAGTGCGCAGAGCAGGACCGACAGCTCACAGACTG 4278
QY 1343 rLysGlnHisArgValSerArg-ArgAlaGlnGlnArgAlaGluSerProGluSerSerA 1363
Db 4279 CGGGCAGAGAAAGGCCAGCTATGCAGAGCAGCTGAGCATGCTGAAGAAAGCGCATGGCCTG 4338
QY 1363 laIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
Db 4339 CT-----GGCAGAGGAGAAACCG 4355
```

Search completed: September 25, 2003, 01:48:10
Job time : 512 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:21:35 ; Search time 769 Seconds

(without alignments)
4882.859 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRTNDGKITPPGVK.....QKGRGRSKRTPSPQPKNV 1391

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Xgapop 10.0 , Ygapext 0.5

Xgapop 6.0 , Ygapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h

-O=/gen2.1/USPTO.spool/US09512581/runat_23092003_163546.5464/app.query.fasta_1.1543

-DB=N_Geneseq_19Jun03 -OPMT=fastac -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0

-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum2 -TRANS=human40.cdi

-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09512581.ecgn.1.1.0.0.runat.23092003.163546.5464 -NCPU=6 -ICPU=3

-NO.MAP -LARGESOURCY -NEG.SCORES=0 -NAIT -DSPBLOCK=100 -LONGLOG

-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N_Geneseq_19Jun03:*

2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7193	100.0	5271	21	AAA28051	Human androgen shu
2	7193	100.0	5355	21	AAA28052	Human androgen shu
3	7155	99.5	7473	23	ABV22430	Human prostate exp
4	7149	99.4	7473	23	ABV28244	Human prostate exp
5	7148	99.4	7473	23	ABV25469	Human prostate exp
6	4760.5	66.2	6744	25	ABX14056	CDNA encoding huma
7	3047	42.4	5469	24	ABO99271	Human coding seque
8	2932.5	40.8	3957	21	AAA47423	Sequence encoding
9	1986.5	27.6	3920	23	ABL22105	Drosophila melanog
10	1729.5	24.0	6990	23	ABL22104	Drosophila melanog
11	1409.5	19.6	2496	24	ABK35358	Human cDNA encodin
12	967	13.4	1104	24	ABK35358	Euchromosome fragm
13	912.5	12.7	7943	22	AAK85948	Human immune/haema
14	864	12.0	714	22	AAK85948	Human full-length
15	762.5	10.6	1217	21	AAK85948	Lung cancer associ
16	748.5	10.4	2295	22	AAK85948	Human secreted pro
17	706.5	9.8	2212	22	AAH17132	Human cDNA sequenc
18	698	9.7	4239	20	AAV87629	EST clone DY17. H
19	645	9.0	772	22	AAH03894	Human gene express
20	609	8.5	738	20	AAZ15259	Human cDNA clone (
21	591	8.2	461	24	ABL89669	Human polynucleoti
22	513	7.1	351	21	AAZ42861	Human 5' EST Isola
23	501	7.0	295	21	AAK02817	Human secreted pro
24	375.5	5.2	1489	22	AAH18066	Human cDNA sequenc
25	370	5.1	785	22	AAH07766	Human cDNA clone (
26	357	5.0	227	24	ABO94421	Tumour suppression
27	349	4.9	530	21	AAZ80598	Human colon cancer
28	276	3.8	433	22	AAK48899	Human cortig polyn
29	248.5	3.5	9361	23	ABV25498	Human prostate exp
30	235	3.3	20448	23	ABL19989	Drosophila melanog
31	232	3.2	8493	25	ACA03930	CDNA downregulated
32	232	3.2	8527	22	AAI60039	Human polynucleoti
33	231.5	3.2	9274	24	ABK13224	DNA encoding novel
34	231	3.2	8503	22	ABL18253	Human polynucleoti
35	230	3.2	12777	23	ABL28641	Drosophila melanog
36	229	3.2	7992	24	ABK32895	DNA encoding novel
37	229	3.2	8083	23	AAK70692	Human coding sequ
38	228	3.2	7034	22	AAH19498	Murine coding sequ
39	225	3.1	5334	19	AAK14598	H. pylori GHP0 175
40	223	3.1	557	24	ABK35631	Human eosinophil-m
41	222.5	3.1	11116	22	AAI59270	Human polynucleoti
42	220.5	3.1	7713	22	AAK51958	Human polynucleoti
43	216.5	3.0	4534	23	ABV22145	Human prostate exp
44	216.5	3.0	4534	23	ABV27985	Human prostate exp
45	216.5	3.0	19303	23	ABL22098	Drosophila melanog

ALIGNMENTS

RESULT 1	
ID	AAA28051
AAA28051	standard; cDNA: 5271 BP.
XX	AAA28051;
AC	
XX	
DT	01-DEC-2000 (first entry)
XX	
DE	Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX	
KW	Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW	chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW	diagnosis; treatment; cytostatic; human; ss.
XX	
OS	Homo sapiens.
XX	
FX	Key Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

FT	exon	1..46
FT		/*tag= a
FT		/number= 1
FT	exon	47..173
FT		/*tag= b
FT		/number= 2
FT	CDS	66..4241
FT		/*tag= c
FT		/product= "AS3"
FT		/note= "Androgen shutof gene 3 protein, the CDS is
FT		specifically claimed as SEQ ID #3"
FT	exon	174..377
FT		/*tag= d
FT		/number= 3
FT	exon	378..464
FT		/*tag= e
FT		/number= 4
FT	exon	465..562
FT		/*tag= f
FT		/number= 5
FT	exon	563..689
FT		/*tag= e
FT		/number= 6
FT	exon	690..854
FT		/*tag= f
FT		/number= 7
FT	exon	855..911
FT		/*tag= g
FT		/number= 8
FT	exon	912..1027
FT		/*tag= h
FT		/number= 9
FT	exon	1028..1122
FT		/*tag= i
FT		/number= 10
FT	exon	1123..1268
FT		/*tag= j
FT		/number= 11
FT	exon	1269..1420
FT		/*tag= k
FT		/number= 12
FT	exon	1421..1534
FT		/*tag= l
FT		/number= 13
FT	exon	1535..1616
FT		/*tag= m
FT		/number= 14
FT	exon	1617..1665
FT		/*tag= n
FT		/number= 15
FT	exon	1666..1805
FT		/*tag= o
FT		/number= 16
FT	exon	1806..1921
FT		/*tag= p
FT		/number= 17
FT	exon	1922..2027
FT		/*tag= q
FT		/number= 18
FT	exon	2028..2188
FT		/*tag= r
FT		/number= 19
FT	exon	2189..2312
FT		/*tag= s
FT		/number= 20
FT	exon	2313..2471
FT		/*tag= t
FT		/number= 21
FT	exon	2472..2540
FT		/*tag= u
FT		/number= 22
FT	exon	2541..2677
FT		/*tag= v

FT	/number= 23
FT	2678..2801
FT	/*tag= w
FT	/number= 24
FT	2802..3006
FT	/*tag= x
FT	/number= 25
FT	3007..3121
FT	/*tag= y
FT	/number= 26
FT	3122..3254
FT	/*tag= z
FT	/number= 27
FT	3255..3374
FT	/*tag= aa
FT	/number= 28
FT	3375..3437
FT	/*tag= ab
FT	/number= 29
FT	3438..3583
FT	/*tag= ac
FT	/number= 30
FT	3584..3689
FT	/*tag= ad
FT	/number= 31
FT	3690..4129
FT	/*tag= ae
FT	/number= 32
FT	4130..4354
FT	/*tag= af
FT	/number= 33
FT	4320..4324
FT	/*tag= ag
FT	/note= "Destabilizing signal"
FT	4355..5253
FT	/*tag= ah
FT	/number= 34
FT	5228..5233
FT	/*tag= ai
FT	5249..5253
FT	/*tag= aj
FT	/note= "Poly adenylation cleavage site"
XX	
XX	WO200050454-A1.
XX	
XX	31-AUG-2000.
XX	
XX	24-FEB-2000; 2000WO-US04732.
XX	
XX	24-FEB-1999; 99US-0121461.
XX	
XX	(TUFTS) COLLEGE.
XX	
XX	Soto AM, Sonnenschein C, Geck P, Szelei J;
XX	
XX	WPI: 2000-565451/52.
XX	P-PSDB; AAY94702.
XX	
XX	New human androgen-induced tumor suppressor cDNA sequence termed
XX	'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
XX	diagnosis and treatment of prostate cancer -
XX	
XX	Claim 1; Fig 1; 152pp; English.
XX	
XX	This invention relates to a human androgen-induced tumour suppressor cDNA
XX	sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
XX	on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
XX	proliferation and use as a marker for the efficient diagnosis and
XX	treatment of prostate cancer. The invention includes AS3 cDNA and protein
XX	sequences, a vector comprising the cDNA sequence, a host cell transfected
XX	with the expression vector, and a method for producing an AS3 polypeptide
XX	comprising culturing the transfected cells. AS3 has cyostatic activity,
XX	and acts to suppress cell proliferation. The AS3 gene is useful as a
XX	

CC marker for the efficient diagnosis and treatment of prostate cancer. The
CC AS3 nucleic acid molecule can be used as a source of antisense agents for
CC sequence specific inhibition of gene expression. The AS3 protein may be
CC used in the treatment of disorders caused by aberrant modification or
CC mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC or aberrant post-translational modification of the AS3 protein. This
CC sequence represents the human AS3 cDNA sequence.

XX
SQ Sequence 5271 BP: 1782 A; 944 C; 1066 G; 1479 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	5271
Score:	7193.00	Matches:	1391
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-512-581b-2 (1-1391) x AAA28051 (1-5271)

QY	1	MetaLHISerLysThrArGThrAsnAspGlyLysIleThr-TyrProGluValLys	20
DB	66	ATGGCTCATCTCAAAACATGACCAATGATGGAATAATTACATCCGCTGGGCTCAAG	125
QY	21	GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys	40
DB	126	GAATATCATGATTAATATCTTAAGAGGATGGTGAAGCATTAAGTGTGTGAAA	185
QY	41	ThrPheMetAspMetAspGlnAspSerGluGluGlyLysGluLeuTyrLeuAsnLeuAla	60
DB	186	ACTTTTATGATATGACGACGACGACTCTGMAAGAAAGACCTTATTAACCTAGCT	245
QY	61	LeuHISLeuAlaSerAspPhePheLeuLysHISProGlyLysAspValArgLeuLeuVal	80
DB	246	TTACATCTTCTTCAGATTTTCTCAAGCATCTCGTAAAGATGTTGCTTACTGTA	305
QY	81	AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro	100
DB	306	GCCTCGCTGCTTGGATATTTTTCAGGATTTATGCTCGAAGCTCCTTACACATCCCT	365
QY	101	AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGlnAspThr	120
DB	366	GATTAACATAAAGATATATTTATGTTTAAACAGACAGTGAAGGGCTAGAGGATACA	425
QY	121	LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlnAsnIleAlaThrValLysSer	140
DB	426	AAGACCCACATCTCAATAGGATTTTATTTACTTGAGACATCTGGCTGCAAGTCA	485
QY	141	TyrAsnIleCysPheGluLeuGlnAspSerAsnGluIlePheThrGlnLeuTyrArgThr	160
DB	486	TATACATATATGCTTGGAGTTAGAAATAGCATGAATTTTCAACCAGCTATACAGAAC	545
QY	161	LeuPheSerValIleAsnAsnGlnIHisAsnGlnLysValHisMetHisMetValAspLeu	180
DB	546	TTATTTTCACTTATTAACATGCGCACATCAAGAAAGTCCATATGACATGCTAGACCTT	605
QY	181	MetSerSerIleIleCysGluGlnLysPheThrValSerGlnGluLeuLeuAspThrValLeu	200
DB	606	ATGACCTCTATTTATTTGTAAGGTAATAGTGTCTCAGAGACCTTTTGATACGGTTTAA	665
QY	201	ValAsnLeuValProAlaHISLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla	220
DB	666	GTAATCTGTGATCGCTCATAGAATTAACAAGCAAGCATATGATTTGGCAAAAGCT	725
QY	221	LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal	240
DB	726	TTACGAAAGAGACAGCTCAACGCTATGAGCCATATATATACCACTTTTATATCAGCTT	785
QY	241	LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu	260
DB	786	CTGATGCTTTGGAAAACATCTATCAAGCATTTGTGTGAGACATGTTCTTTACTTAATTTTG	845
QY	261	GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlnPheLys	280

DB	846	GAGCTCTACATATGATGATGATTCATTCGCTCTCTGTTTACCCAGCTGAATTA	905
QY	281	LeuLysSerAsnAspAsnGlnLysArgLeuGlnValValLysLeuLeuAlaLysMetPhe	300
DB	906	TTAAAGAGCAATGAATATAGAGAGCCCTACAGATGTTTAACTATGGCAAAAATGTTT	965
QY	301	GlyAlaLysAspSerGluLeuValAserGlnAsnLysProLeuTrpGlnCysTyrLeuGly	320
DB	966	GGGGCAAGGATTCACAAATTTGGCTTCTCAAAACACCCCTTTGGAGTGCTACTGGCC	1025
QY	321	ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys	340
DB	1026	AGGTTTAATGATATCCATGATACCAATCCCGCTGGAATGTGTGAAATTTCTGACCATGT	1085
QY	341	LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis	360
DB	1086	CTCATGAACCATCTGATTTTACCAAAAGACTTAAACAGACTATCTTAAAGTGAAGTCACAT	1145
QY	361	AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys	380
DB	1146	GACCTCGAGGAAGCTATTGACATGATGTTATTTGTCTCAATAGTACCTGCTAAAG	1205
QY	381	AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys	400
DB	1206	GATATTCCTGCTGATCATATCATCTTACTTAATTTTGTGAGAGAGACATTAACAGAAA	1265
QY	401	ArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla	420
DB	1266	CGATGGAGAGTACGCAAAAGACCAATGAGGAGACTGCCCAATTTATTAAGAAATATGCT	1325
QY	421	LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu	440
DB	1326	TTACAGTCAGCACTGGAAAGATGCTGCAAAACAGATGCAATGCAATCAAGCAATATG	1385
QY	441	LeuHISLeuTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla	460
DB	1386	CTACATATATATATCAAAATAGTATGATGATCGACTGTTGTAACGAGATCTTGTCT	1445
QY	461	GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrLeu	480
DB	1446	CATACATNTGGTTCCTCAAAATTTAGAAATACAGAAACGATGAAATGCTTAAATTAATCTTG	1505
QY	481	TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn	500
DB	1506	TATGCCACCTGATTTAAATGCTGTGAAGCATTTGAATGAATGGGAAATGTCAAAT	1565
QY	501	LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla	520
DB	1566	CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAAGCAACCCAAACAGATGCC	1625
QY	521	SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
DB	1626	AGTGTCAAGCCATATTTTCAAAAGGATTAATTAACAAGAAATTTACCTATCTCTGCT	1685
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg	560
DB	1686	AAGGCTCAGGATTTCAATGAGAAATTCACACAGGTGTTGAAGATGATAGAAAATTAAGA	1745
QY	561	LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluLysCysVal	580
DB	1746	AAGCAGTTTGAAGTACTTGTAGTCCAAACATGCTCTGCAACAGAGCTGAAGGTTGTGTG	1805
QY	581	ArgGluIleThrLysLysLeuGluLysAsnProLysGlnProThrAsnProPheLeuGluMet	600
DB	1806	CGTGAAATTAACCTAAGAGTTGGCAACCCCAAAACACCTTACAAATCTCTTGGAAGG	1865
QY	601	IleLysPheLeuLeuGlnArgIleAlaProValHisIleAspThrGluSerIleSerAla	620
DB	1866	ATCAAGTTTCTTGTGAGAGGATAGACCTGTGACATATGATACCAATCTATCAGTGT	1925
QY	621	LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal	640

Dd 1926 CTATTATAACAGTAGAACAAATCAATAGATGGAACACAGATGATGAAGATGAGGCTGTT 1985
Qy 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuValLeuSerPheThrHis 660
Dd 1986 CCAACTATATCAGGCATCAGAGCAGGCTTGAACCTGCTTAAGGACTCTCTATTACACAT 2045
Qy 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Dd 2046 CCATCTCATCTTCTCTGCGAAGCAATTGCAATCATTAAGCTGCTGCTGCAAAATGAT 2105
Qy 681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Dd 2106 GATCAAAAGATGACAGAGAGCTGCACTCAAAATTTTCAAAAAACAGAGCAAAATTGAA 2165
Qy 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLysLys 720
Dd 2166 GAGATTTTCCACACATCAGATCGCTTGCTGCTCTGTTTACATCCAAATCTAAAAA 2225
Qy 721 GlyProDArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
Dd 2226 GGAACCCCGCTCAAGCCAAATATGCCATTCATGTATCCATGCGATATTTTCTAGTAAA 2285
Qy 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLysAspProSerAsnLeu 760
Dd 2286 GAGACCCAGTTTGACAGATATTGAGCCTCTGCATAGAGCCTAGATCCAGAACCTCG 2345
Qy 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Dd 2346 GAACATCTCATTAACACCATGTTGTTACTATTGGTCATATTGCTCTCCCTGACCGATCAA 2405
Qy 781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Dd 2406 TTGGCTCTCTCTGGAAATCTGGGTAGCTAATTTCATTGGAAGATCTTCTCATGAT 2465
Qy 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
Dd 2466 GATGGCTTCCAGGAGAAAAAGACAACTAATCTGGGTTCCAGATGGAAGATATCTCCT 2525
Qy 821 GluThrMetValLysIleGlnAlaIleLysMetLeuValArgTrpLeuLeuGlyMetLys 840
Dd 2526 GAGCAATATGTCAAAATTCAGGCTTTAAATGATGTGTCGATGCTTGAATGAA 2585
Qy 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
Dd 2586 AATATATACAGTAAATACAGAACTTCTAACCCTTAAGATTGCTAACACAAATTTCCATAGT 2645
Qy 861 AspIleAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Dd 2646 GATGAGACTTGCACAGAACAGGGAATTTAGTAAACCAGATATGTCACTGTCAGACTT 2705
Qy 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnIleProCysTyrHisGluIleLeuThr 900
Dd 2706 GCTCTCTGGAGTCTATTGTGTAAAGCTGGCACAAAGCCCTGTACCATGAAATCATCA 2765
Qy 901 LeuGluGlnTrpGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Dd 2766 TTAGAACAAATATACAGTATGTGCATATGCTAACAGATGATCTATCAAGTAGAGAA 2825
Qy 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnTyrMet 940
Dd 2826 GTGTTTGGCCGAAAGACTTCCAAAGGCTTCCGCTTACGCGCTTCCACTTGAGTATATG 2885
Qy 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Dd 2886 GCATATCTGTGCCCTTTGTGCAAAAGATCCCTTAAGAGAGAGAGAGGCTCATGTAGGCAA 2945
Qy 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
Dd 2946 TGTTTGGTGAATAATATATATATGTAAGCGGAGATATGTGAAGACAGCATGCGCTGTAGT 3005
Qy 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Dd 3006 GAAAAATTAATTTCTCTTCTACAGAGATATGTGTCTCATATTAACAAATTCACCTTTTGGCA 3065

Qy 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGlyLys 1020
Dd 3066 CATACCCAGATTTATGTCMAAGTACAGGATATTTCACAACTTAAGATGTTAAAGATGT 3125
Qy 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Dd 3126 CTTTGGTTTGTTCGAAATATTAATGCTTAAATATGAAATTAACATCCAGCTTTTATTC 3185
Qy 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Dd 3186 AGAAGATGTGAGAAATATTAAACAACAAGATGCCCAAGAGCAGATGATGCAAAA 3245
Qy 1061 MetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSer 1080
Dd 3246 ATGAATGAATAAACTGTACAGTGTGTGATGTTCCATGAATATCATCATGTCAAAGAT 3305
Qy 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
Dd 3306 ACTACATACAGTTTGGAAATCTCCCTAAAGACCCGGTACTACAGCTCTTCTTCACACAA 3365
Qy 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Dd 3366 CTTGACAAAGATTTCACTAACACCAAAATATATGCTCTCTGAATGAATCATTTTTC 3425
Qy 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Dd 3426 ACTCCTGGAATAAACCCTAAACAAACCAATGTTCTAGACCTGTAAACAGCCACTTCATCA 3485
Qy 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Dd 3486 GCAGGCAAGCATATCTCAGACCAATATCATCAGAAATGGAACCTGTAACCAATGCAAGCAGC 3545
Qy 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Dd 3546 AGCTCAATCCAACTCTCCCTCGAAGAATAAAGGGAGCGTGTATGTTCTGAAATGAT 3605
Qy 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Dd 3606 CACAGTGAATAATGAAATTCACAAATCTTCCACTTGGCCGGGAAAAAAGACAAAG 3665
Qy 1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
Dd 3666 AGAGACAGCTGTGATTTGTGAAGCTGTGAATTTGAGAAAGCTACAGGCAGAGAAAAAAGC 3725
Qy 1221 ProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
Dd 3726 CCCGTCAAGAAAGAGAGGAATAATTAGTATGATGACTGACTGAAGTTGCTACAGAA 3785
Qy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
Dd 3786 CAGAAACCTTAAGGACAGTACAGCAAGTCGGAAGAGGCCCTACGGCTTCGAATTCGAT 3845
Qy 1261 GlnGlnGlnTrpProGluGluLysArgLeuLysGlnAspIleLeuGlnIleAsnGluAspGlu 1280
Dd 3846 GAACAGCAGTGGCTTGAGGAAGAGAGGCTCAAAAGATATATTAGAAATGAAGATGA 3905
Qy 1281 GluAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
Dd 3906 CAGATATGTCCGCCAAAAAAGGTAAAGAGCGGACACCAAAACCTCTTGCGAGAGT 3965
Qy 1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
Dd 3966 ACACCAAAAGAGAGCCAAACAAATGAAACTCTTAAGAAAGAGAACAAAAAATCTGGA 4025
Qy 1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
Dd 4026 CCTCCAGCACAGAGGAGAGAGAGAAAGAAAGAAACAAATGTGAATATACGAACACAG 4085
Qy 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Dd 4086 TCCAAAGCCAAACAGCACCGAGTGTCAAGAGAGACACAGCAGAGACAGAAATCTCTGAA 4145

PR 24-FEB-1999; 9905-0121461.
 XX (TUFT) TUFTS COLLEGE.
 PA Soto AM, Sonnenschein C, Geck P, Szelei J;
 PI WPI; 2000-565451/52.
 XX P-PSDB; AAY94702.
 DR
 XX
 PT New human androgen-induced tumor suppressor cDNA sequence termed
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
 PT diagnosis and treatment of prostate cancer -
 XX
 PS Example 4; Fig 6; 152pp; English.
 XX
 CC This invention relates to a human androgen-induced tumour suppressor
 CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in
 CC inhibiting cell proliferation and use as a marker for the efficient
 CC diagnosis and treatment of prostate cancer. The invention includes AS3
 CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host
 CC cell transfected with the expression vector, and a method for producing
 CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
 CC cytosolic activity, and acts to suppress cell proliferation. The AS3
 CC gene is useful as a marker for the efficient diagnosis and treatment of
 CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
 CC antisense agents for sequence specific modulation of gene expression. The
 CC AS3 protein may be used in the treatment of disorders caused by aberrant
 CC modification or mutation of a gene encoding an AS3 protein, misregulation
 CC of the AS3 gene or aberrant post-translational modification of the AS3
 CC protein. This sequence represents the human AS3 cDNA sequence with an
 CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
 CC compared with the claimed AS3 cDNA sequence AAA28051.
 CC
 XX
 SQ Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 5355
 Score: 7193.00 Matches: 1391
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-512-581b-2 (1-1391) x AAA28052 (1-5355)

QY 1 Metalh1sSerlystrArgThrAsnAspGlyLysIleThrTyrrProProGlyVallys 20
 DB 150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAATAATTCATATCCGCGCTGGGGTCAAG 209
 QY 21 GluileSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValVallys 40
 DB 210 GAATATCATGATAAATATCTAAAGAGAGATGGTGAGCATTAAGATGGTGTGATAA 269
 QY 41 ThPhemetAspMetAspGlnAspSerGluGluLysGluLeuTyrrLeuAsnLeuAla 60
 DB 270 ACTTTTGTGATATGACACGAGACTCTGAAGAAAGAGAGCGCTTTTAAACCTAGCT 329
 QY 61 LeuH1sLeuAlaSerAspPhePheLeuLysH1sProGlyLysAspValArgLeuLeuVal 80
 DB 330 TTACATTTGCTTCACAGTTTTTTCTCAACGATCTCTGTAAAGATGCTTCGTTACTGTA 389
 QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrrAlaProGluAlaProTyrrThSerPro 100
 DB 390 GCGCGCTGCTTGCTGATATTTTCAGATTTATCTCTGAGCTCTTACACATCCCT 449
 QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgLysLeuLysGlyLeuGluAspThr 120
 DB 450 GATAAACTAAAGGATATATTATGCTTTATAAACAAGAGCTGAAGGGGCTAGAGATACA 509
 QY 121 LysSerProGlnPheAsnArgTyrrPheTyrrLeuLeuGluAsnIleAlaTTPValLysSer 140
 DB 510 AAGAGCCCAATTCATAGGATTTTATTTACTTGAAGACATTGCTGGGTCAAGTCA 569

QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGluLeuTyrrArgThr 160
 DB 570 TATACATATGCTTGTGATAGAAAGATAGAAATTTTACCCACGCTATACAGAAC 629
 QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGluLysValHisMetHisMetValAspLeu 180
 DB 630 TTATTTTCACTTAATAACAATGGCCCAACAAACAGAAAGTCCATATGACATGATGACCTT 689
 QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGluLeuLeuAspThrValLeu 200
 DB 690 ATGAGCTCTATTTTGTGAGGATGATACGTGTCTGAGAGCTTTTGGATAGGTTTA 749
 QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGluAlaTyrrAspLeuAlaLysAla 220
 DB 750 GTAATTTGGTACTGCTGCTATTAAGAAATTTAAACAGACACATATGATTTGGCAAGGCT 809
 QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrrIleThrThrPhePheAsnGlnVal 240
 DB 810 TTACTGAAGAGGACGCTCAAGCTATGAGCCATATATTCACACTTTTAAATCAGCTT 869
 QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
 DB 870 CTGATGCTTGGGAAACATCTATCAGCGATTTGTACAGCATGCTTTGACTTATTTTG 929
 QY 261 GluLeuTyrrAsnIleAspSerHisLeuLeuLeuSerValLeuProGluLeuGluPheLys 280
 DB 930 GAGCTTACATATTTGATGATGATCATTTTCTGCTCTCTGTTTACCCAGCTGAATTTAA 989
 QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
 DB 990 TTTAAAGAGCATGATGATATGAGGAGCGCTTCAAGTTGTTAACTACTGCGGCAAAATGTTT 1049
 QY 301 GlyValAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrrGlnCysTyrrLeuGly 320
 DB 1050 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAACAAGCCTTTGGCGCTACTTGGCC 1109
 QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
 DB 1110 AGTTTATGATATATTCATGATACCAATCCGCTGGAAATGTGAAATTTGCTAGCATTTGT 1169
 QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrcLysTyrrLeuLysValArgSerHis 360
 DB 1170 CTCATGAGACCATCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGTAGGCTCAT 1229
 QY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrrAlaAlaLysLys 380
 DB 1230 GACCTTAGGAGGCTATTAACATGATGATTTTGTGTCAATAGTTACAGCTGCTAAAG 1289
 QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
 DB 1290 GATATCTCTGCTGCTCATGATGATCACTTAATTTTGTGAGAGAGAACATTAAGACAA 1349
 QY 401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrrLysTyrrAla 420
 DB 1350 CGATGGAGAGTACCCAAAGAAAGCCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1409
 QY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
 DB 1410 TTACAGTACAGCAGCTGGAAGAAAGATGCTGCCAAACAGATACATGATGATCAACAAATTTG 1469
 QY 441 LeuHisIleTyrrTyrrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
 DB 1470 CTACATATATATTTCAAAATAGATATGATGATGATGATGATGTTGTAAGCGATCTTGGCT 1529
 QY 461 GlnTyrrMetValProHisAsnLeuGluThrThcGluArgMetLysCysLeuTyrrTyrrLeu 480
 DB 1530 CAATACATGTGCTCTCCCAATTTTGAAGAACACAAAGGATGGAATGCTTATATTTACTTGG 1589
 QY 481 TyrrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTyrrPlyLysCysGlnAsn 500
 DB 1590 TATGCCACTGCTGATTTTAAATGCTGTGTGAAGCATTTGAATGTGGAATGTCAAAAT 1649

```
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuLeuLeuGlnProLysThrAspAla 520
      |||||
      1650 CTGCTCCGACATCAAGTAAAGATTGGCTTGACTTATTAAGACCCAAACAGATGCC 1709
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
      |||||
      1710 AGTGTCAAGGCATATTTTCAAAAGATGATGTTATTACAAAGAAATTACTGATCTGGG 1769
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg 560
      |||||
      1770 AAGGCTCAGGATTTTATGAAAGAAATTCACACAGGTGTTAGAAAGTATGAGAAATAA 1829
QY 561 LysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnIleGluLysVal 580
      |||||
      1830 AAGCATTTGAAAGTACTTGTAGTCCAAATGCTCTCGAAGCAGGCTGAAGGTGTGG 1889
QY 581 ArgGluIleThrLysLysLeuGlnLysAsnProLysGlnProThrAsnProPheLeuGlnMet 600
      |||||
      1890 CGTGAATATTAAGAAAGTTGGGCAACCCCAACACCTTACAATCTTCCGGAAATG 1949
QY 601 IleLysPheLeuLeuGlnArgIleAlaProValHisIleAspThrGlnSerIleSerAla 620
      |||||
      1950 ATCAAGTTTCTTGTGAGAGGATGACCTGTGCACATGATAGATACCAATCTTACCTGCT 2009
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal 640
      |||||
      2010 CTTATTAAACAAGTGAACAAATCAATAGATGGAACGACGATGATGAATGAGGTGTT 2069
QY 641 ProThrAspGlnAlaIleArgAlaGluLeuGlnLeuLeuValLeuSerPheThrHis 660
      |||||
      2070 CCAACTGATCAAGCATCAGACAGCTGTGAACCTGTGAACCTGTCAATTCATTTACAT 2129
QY 661 ProIleSerPheHisSerAlaGluThrPheGlnSerLeuLeuAlaCysLeuLysMetAsp 680
      |||||
      2130 CCCATCTCTATTCTCATTTCTGCAACATTTGAATCTTACTGCTGTGTGAATAATGAT 2189
QY 681 AspGluLysValAlaGlnAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
      |||||
      2190 GATGAAAAAGTAGCAAGACCTGCACATACAAATTTTCAAAAACACAGGAAGCAAAATTGA 2249
QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
      |||||
      2250 GAGGATTTTCCACACATCAGATCAGCTGCTGCTGTTTACATCACAAATCTTAAAAA 2309
QY 721 GlyProArgGlnAlaLysTyraIleHisCysIleHisAlaIlePheSerLysLys 740
      |||||
      2310 GGACCCCCCGCCAAAGCCAAATATGCCATTTCTGATCCATGCGATATTTTCTAGTAA 2369
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
      |||||
      2370 GAGACCCAGTTTGCACAGATATTGAGCCTTGCATAGAGCCTAGATCCAAAGCAACCTG 2429
QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
      |||||
      2430 GAACATTCATTAACCATTTGGTTACTATTGGTCAATTGCTCTCTGACCTATCA 2489
QY 781 PheAlaAlaProThrLysSerThrValAlaThrPheIleValLysAspLeuLeuMetAsn 800
      |||||
      2490 TTTGCTGCTCTTGGAAATCTTGGGTAGCTACTTTGATTTGAAAGATCTTCTCAAGAT 2549
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuThrValProAspGluGluValSerPro 820
      |||||
      2550 GATCGGCTTCCAGGAAAAAGCACTTAACTTTGGGTTCCAGATGAAGAACTATCTCTCT 2609
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlnLysLys 840
      |||||
      2610 GAGAAATGATGCAAAATTCAGGCTTATAAAAGATGATGCTGATCTTGTGAATGAAA 2669
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
      |||||
      2670 AATATATCAGATTAATCAGAACTTCTACCTTAATATGCTTACAAACATATTTGCAATG 2729
QY 861 AspGluAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
      |||||
      2730 GATGGAGACTTGACAGAACAGGGAAAAATTAATGAACGATATGTCAGCTGTGACACTT 2789
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrrHisGluIleIleThr 900
      |||||
      2790 GCTGCTGGAGAGTCTATTGTGAAAGCTGGCAAGAACCTGTATACATGAATCATACACA 2849
QY 901 LeuGluGlnTrpGlnLeuLeuCysAlaLeuAlaIleAsnAspGluCysTyrrGlnValArgGln 920
      |||||
      2850 TTAGAACATATTCAGCTATGTGCATTTAGCTATCAACGATGAATGCTATCAACCTAAGACA 2909
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnLysTrpMet 940
      |||||
      2910 GTGTTTGCCCAAGAACTTACAAAGGCTTTCCCGTTTACGCGCTTCCACTTGAGTATAG 2969
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
      |||||
      2970 GCAATCTGCGCCTTGTGCAAAAGATCTGTAAAGAGAGAGAAAGACCTCATGCTAGGCAA 3029
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTrpLeuLysGlnHisAlaAlaValSer 980
      |||||
      3030 TGTTTGGTGAATAATTAATGTAAGCGGGAGTATCTGAAACGACATGACCTGTTAGT 3089
QY 981 GluLysLeuLeuSerLeuLeuProGluTrpValValProTyrrThrIleHisLeuLeuAla 1000
      |||||
      3090 GAAAAATTTTGTCTCTTCTTACCAGATATGTTGTTCCATATACAAATTCACCTTTGGCA 3149
QY 1001 HisAspProAspTrpValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
      |||||
      3150 CATGACCCAGATATGTCAAAAGTACAGGATTAATGAACAACCTTAAGATGTTAAAGATGT 3209
QY 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGlnAsnSerHisAlaPheIle 1040
      |||||
      3210 CTTTGGTTTGTCTGGAATATTAATGCTAAATAATGAAATAAACAGTACCTTTTATC 3269
QY 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
      |||||
      3270 AGAAAGATGTGAAATAATTAATTAACAAACAAAGATGCCCAAGACAGATGATCAAAA 3329
QY 1061 MetAsnGluLysLeuTrpThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
      |||||
      3330 ATGAATGAAAAACTGTACCTGTGTGTGTTGGCATTGAATAATATCATGCAAAAGAT 3389
QY 1081 ThrThrTyrrSerLeuGlnLysSerProLysAspProValLeuProAlaArgPheThrGln 1100
      |||||
      3390 ACTACATACAGTTTGGAAATCTCTTAAGACCCGGTACTACAGCTGCTTCTTCACTCA 3449
QY 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrrLeuProProGluMetLysSerPhePhe 1120
      |||||
      3450 CCTGACAAAGAAATTTTCAGTAAACACCAAAATTAATATGCGCTTCCGAATGAAATCATTTTTC 3509
QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
      |||||
      3510 ACTCCTGAAAAACCTTAAACCAACCAATGTTCTAGAGCGGTTTAACAAAGCATTTTCAACA 3569
QY 1141 AlaGlyLysGlnSerGlnThrThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
      |||||
      3570 CGAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAACCTGTAAAGCAATCAAGCAC 3629
QY 1161 SerSerAsnProSerSerProGlyIleLysGlyArgLeuAspSerSerGluMetAsp 1180
      |||||
      3630 ACTCAAAATCCAAAGCTCTCCGTAAGAAATAAAGGGAGGCTTGATAGTTCTGAATGAT 3689
QY 1181 HisSerGluAsnGluAspTyrrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
      |||||
      3690 CACACTGAAAAAGAAAGATTACCAATGCTTACCTTGGCCGGGAAAAAAAGTCAAG 3749
QY 1201 ArgAspAspSerAspLeuValArgSerGlnLeuGlnLysProArgGlyArgLysLysThr 1220
      |||||
      3750 AGAGCGACTGCTGATCTTGTAAAGTCTGAATTTGGAAAGCCTTAGAGGCAAGAAAAAAGC 3809
QY 1221 ProValThrGluGlnGluLysLysLeuLysMetAspAspLeuThrLysLeuValGlnGlu 1240
      |||||
```

Db 3810 CCCCTCACAGACGAGGAGAAATTAGCATGATACCTTGACTAAAGTTGGTACAGAA 3869
QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
Db 3870 CAGAAACCTTAAGGCACTACGCGAAGTCGGAAGAGCCATACGGCTTCAGATCTGAT 3929
QY 1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
Db 3930 GAACACGACGTGCCTGAGAAAGAGGCTCAAGAGATATATATGAAAATGAAAGATGAA 3989
QY 1281 GluAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGly 1300
Db 3990 CAGATATGTCGCCAAAAAGGATTAAGAGCCGACCAACCACTCTTGTTGAGAGT 4049
QY 1301 ThrProLysGluGluProTrpMetLysThrSerLysGlySerLysLysSerGly 1320
Db 4050 ACACCAAAAGAGAGCCACACATGAAACTCTTAAAAAGAGAACCAAAAAATCTGGA 4109
QY 1321 ProProAlaProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
Db 4110 CCTCCACGACGAG 4169
QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlu 1360
Db 4170 TCCAAAGCAAAACGACCCGAGTGTCAAGGAGACACAGAGAGAGAGAGAGAGAGAGAG 4229
QY 1361 SerSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4230 TCTAGTGCATTTGATTCACACAGCTCCACACACAGAAAGAGAGAGAGAGAGAGAGAG 4289
QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4290 ACGCATCACCATCATCAACCAAAAAAATGTG 4322
RESULT 3
ABV22430
ID ABV22430 standard; cDNA: 7473 BP.
XX
XX ABV22430:
AC
XX 13-SEP-2002 (first entry)
DT
XX Human prostate expression marker cDNA 22421.
DE
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001WO-US05171.
PE
XX
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 3901-3902; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;
Alignment Scores:
Pred. No.: 0 Length: 7473
Score: 7155.00 Matches: 1387
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 4
Query Match: 99.47% Indels: 0
DB: 23 Gaps: 0
US-09-512-581b-2 (1-1391) x ABV22430 (1-7473)
QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrYrProProGlyValLys 20
Db 158 ATGGCTCATTCAAAGACTAGACCAATGATGAAATAATCATATCCGCTGGGGTCAAG 217
QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValLys 40
Db 218 GAATATACAGTAAATATCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuLysGluLeuLys 60
Db 278 ACTTTTATGATATGACACGAGACTGTGAAGAAAGAGAGCTTTATTAACCTGAGCT 337
QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
Db 338 TTACATCTTCCTTCAGATTTTCTTCACAGATCCATTAAGAGTTCGCTTACGTGTA 397
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
Db 398 GCGTGTGCGCTGCTGATATTTTCAGATTTTATGCTGAGAGTCTTACACATCCCT 457
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
Db 458 GATTAACCTAAAGCATATATTTATTTATTAACAAGACAGTGAAGGGGCTAAGAGATACA 517
QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuLysAsnIleAlaIleTyrLysSer 140
Db 518 AAGAGCCACACATTCATAGGTATTTTATTTACTTGAAGACATGCTGGGTCAAGTCA 577
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
Db 578 TATTAACATATGCTTTGAGTTAAGAAGATAGCATTAATTTTCACCCGCTATACAGAAC 637
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
Db 638 TTATTTTCAGTTATTAACAATAGGCCACACATCAGAAATCCATATGACATGTAGACCTT 697
QY 181 MetSerSerIleIleCysGluGluArgPheValSerGlnGluLeuLeuAspThrValLeu 200
Db 698 ATGAGCTCTATTTATTTTGAAGGTGATACAGTGTCTGAGAGCTTTTGGATACGTTTGA 757
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
Db 758 GTAAATCTGTAAGTCTCATTAAGAAITTTAAACAGACAGACATATGATTTGGCAAGGCT 817

Db	1898	CGTCAAAATTAACCTAAGAGTGGGCAACCCCAACAGCCTACAAATTCCTTCCCTGCAATG	1957
Qy	601	IIlelySpheuleuengluargIIealaProValHisIIleSprngIuserIIleSerAla	620
Db	1958	ATCAAGTTTCTCTGGAGAGATAGCACTGTGTGCACATAGATACCGAATCTATCAAGTCT	2017
Qy	621	LeuIIlelySGInValAsnIysSerIIeasPglYThrAlaAspAspGluAspGluIVal	640
Db	2018	CTTATTTAACAAGTCGAACCAATTCATATAGATGGAACACAGATGATGAAGATGAGAGTGT	2077
Qy	641	ProThrAspGluAlaIIeargAlaGluIleuGluIleuLeuIysValIleuSerPheThrHis	660
Db	2078	CCAACGTATCAAGCCATCAGCAGCGTCTTGAATGCTTAAGTACACTCTATTACACAT	2137
Qy	661	ProIIleSerPheHisSerAlaGluThrPheGluSerIleuLeuAlaCylSerIlyMetAsp	680
Db	2138	CCCATCTCATTTCTCTGTGTAACCAATTTGAATCATCTACGCGCTGTCTGTAAATGGAT	2197
Qy	681	AspGluIysValAlaGluAlaIleLeuGluIIlePheIysAsnThrGlySerIysIIeGlu	700
Db	2198	GATGAAAAAAGTAGAGAGAAGCTGCACCTACAAATTTTCAAAAAACACAGAGCAAAATTTGAA	2257
Qy	701	GluAspPheProHisIleargSerAlaIleuIleuProValIleuHisIlySerIysIys	720
Db	2258	GAGATATTTTCCACACATCAGATCAGCCTGTGCTCTGTTTACATCAACAAATCTAAAAA	2317
Qy	721	GlyProProArgGluAlaIysTyValAlaIleHisCylIleHisAlaIIlePheSerSerIys	740
Db	2318	GGACCCCCCGCTCAAGCCAAATATAGCCATTCATGTATCCATGGCAATATTTTCTAGTAA	2377
Qy	741	GluThrGlnPheAlaGluIIlePheGluProLeuHisIysSerIleAspProSerAsnIleu	760
Db	2378	GAGACCCAGTTTGGACAGATATTTTGAAGCTCTGCATTAAGACCTTAGATCCAGCAACCTG	2437
Qy	761	GluHisIleuIleThrProIleuValThrIIeGlyHisIIealaIleuAlaIleProAspGln	780
Db	2438	GAACTCTCATACAAACCATGTGTACATATGGTCATATGGCTCTCTCCCTGGCACTGATCAA	2497
Qy	781	PheAlaIleIleProIlySerSerThrValAlaThrPheIleValIysAspIleuIleMetAsn	800
Db	2498	TTTCTCTCTCTTGGAAATCTTGGTAGCTACTTTCATGTGGAAGATCTTCTCATGAT	2557
Qy	801	AspArgIleuProGlyIysIysThrThrIysIleuThrValProAspGluIleValSerPro	820
Db	2558	GATGGCTTCCAGGGAAAAAGACAACCTAACCTTTGGGTTCCAGATGGAAGAAGTATCTCCT	2617
Qy	821	GluThrMetValIysIIeGluAlaIIeIysMetMetValArgThrIleuGluIlyMetIys	840
Db	2618	GAGACAAATGGCAAAATTCACAGCTATTAATAATGATGTGTCATGGCTACCTGGAATGAA	2677
Qy	841	AsnAsnHisSerIysSerGlyThrSerThrIleuArgIleuIleThrThrIIeIleHisSer	860
Db	2678	AATGATATCACAGTAATACAGAACTTCTACCTTTAAGATTTGCTAACACAAATATTCATAGT	2737
Qy	861	AspGlyAspIleuThrGluGluGlyIlyIleSerIysProAspMetSerArgIleuArgIleu	880
Db	2738	GATGGAGACTTGCACAGAACAGGGAAAAATTAAGTAACACAGATATGTACAGCTTGACACTT	2797
Qy	881	AlaAlaIlySerAlaIIeValIysIleuAlaGluIleuProCysTyThrIlyGluIleIleThr	900
Db	2798	GCCTCTGGAGTGTATTATTTGGAACCTGGCAACAACACCTGTTTACATGAATCATACACA	2857
Qy	901	LeuGluIleuThrGluIleuCysAlaIleuAlaIIeAsnAspGluCylSerTyGlnValArgGln	920
Db	2858	TTTAAACAATATCAGCTATGTGCTATTTAGCTATCACACATGATGATCTTCAAGTAAGCAAA	2917
Qy	921	ValPheAlaGluIlyIleuHisIlyGlyIleuSerArgIleuArgIleuProIleuGluIlyMet	940
Db	2918	GTTGTTTCCCGCAAAACTTCAACAAAGGCTTTCCGTTTACGGCTTCCACTTGAGATATATG	2977
Qy	941	AlaIleCysAlaIleuCysAlaIysAspProValIysGluIlyArgAlaHisIleAlaArgGln	960

Dd	2978	CCAAATCTGTGGCCCTTTGGTGCACAAAAGATCTCTAAAGGAGAGAAAGGCTCATGCTAGGCAA	3037
Qy	961	CysLeuValIlySasNIleasnValIArgArgIuTyLeuIySgInHIsAlaIalaValSer	980
Dd	3038	TGTTTGGTGAATAAAATATAATGAAGGGGAGATATCGAAGCAGCAAGCAGCTGTAGT	3097
Qy	981	GIuTySleuLeuSerLeuLeuProGIuTyTyValValProTyThrIleHIsleuLeuIa	1000
Dd	3098	GAAAAATATATGTCTCTTACACAGAGTATGTTGTTCATATTCATATTCACCTTTTGSCA	3157
Qy	1001	HisAspProAspTyTyValIySValGIasnPIleGIuLeuIySAspValIySGLuCs	1020
Dd	3158	CATGACCCGAGATTATGATCAAGTACAGATATTGACCAACTTAAAGATGTAAAGAAAGT	3217
Qy	1021	LeuTyPheValIleuGIuIleIleuMetIAlaIySasnGIuIySasnSerHIsAlaPheIle	1040
Dd	3218	CTTGCGTTGTCTTCGGAATATTAATAGCTTAAATGAAATTAACAGTCACGCTTTTATC	3277
Qy	1041	ArgIySMeTValGIuAsnIleIySGLInThrIySAspAlaGIuIyProAspAspAlaIyS	1060
Dd	3278	AGAAAAGATGTAGAAAATATTAACCAACAAAGATGCCCAAGGCCAGATGATGCANAA	3337
Qy	1061	MetasnGIuIySLeuTyTyThrValIyCysAspValAlaMetasnIleIleMetSerIySer	1080
Dd	3338	ATGATATGAATAACCTGTACACGTGTGATGATTTGCCAGATATCATCATCTCAAGAGT	3397
Qy	1081	ThrTyTySerLeuGIuSerProIySAspProValIeuProAlaArgPhePheThrGIu	1100
Dd	3398	ACTATATACATTTGGATCTCTCTTAACACCCGGACTACTACAGCTCTTTCTCACTCA	3457
Qy	1101	ProAspIySAsnPheSerAsnThrIySAsnTyIleuProProGIuMetIySerPhePhe	1120
Dd	3458	CCTGCACAGAAATTTTCAGTACACCAAAAATATCTGCCTCTCGAAATGAATTCATTTTTC	3517
Qy	1121	ThrProGIuIySProIySThrThAsnValIeuGIuIalaValaIySProIeuSerSer	1140
Dd	3518	ACTCCTGGAAAAACCTAAAAACAACCAATTTCTTAGAGCTGTTAACACCCACTTTCATCA	3577
Qy	1141	AlaGIuIySGLInSerGIuThrIySAspSerArgMetGIuThrValSerAsnAlaSerSer	1160
Dd	3578	GCAGCGCAAGCATCTCAACACCAAAATCAACAGAAATGGAACTGTAAAGCAATGCAAGC	3637
Qy	1161	SerSerAsnProSerSerProGIuYArgIleIySGLArgIeuAspSerSerGIuMetAsp	1180
Dd	3638	AGCTCAATTCACAAGCTCTCTCGGAAGATTAAGGGAGGCTTGATAGTTCGAAATGAT	3697
Qy	1181	HisSerGIuAsnGIuAspTyTyThrMetSerSerProIeuProGIuIySLeuSerAspIyS	1200
Dd	3698	CACAGTCAAAATGAAGATTACACAATGCTCTCACCTTGCAGGGGAAAAAAGTGACNAG	3757
Qy	1201	ArgAspAspSerAspIeuValArgSerGIuIeuGIuIySProArgGIuArgIyIySthr	1220
Dd	3758	AGAGACGACTGTATCTTAAAGGCTGAATTTGGAGAGCCTTAGAGCGAGGAAAAAAGCG	3817
Qy	1221	ProValThrGIuGIuGIuGIuIySLeuGIuIyMetAspAspIeuThrIySLeuValGIuGIu	1240
Dd	3818	CCCGTCACAGAAACAGAGAGAAATTAAGTATGATAGACTTGACATAATGTTGTCACAGAA	3877
Qy	1241	GIuIySProIySGLySerGIuArgSerArgIySArgGIuHIsThrAlaSerGIuSerAsp	1260
Dd	3878	CAGAAACCTTAAGGAGTACACCGAAGTTCGAAAAAGAGCCCATACGCGCTTCAGAAATCTGAT	3937
Qy	1261	GIuGIuGIuIntProGIuGIuIySArgIeuIySGLuAspIleIeuGIuAsnGIuAspGIu	1280
Dd	3938	GAACACGCGTGGCTGAGGAAAAAGAGGCTCAAAAGACATATATTAAGAAAAAGAAAGTGA	3997
Qy	1281	GIuAsnSerProProIySLeuGIuIySArgIyIyArgProProIySProIeuGIuGIuGIu	1300
Dd	3998	CAGATATGTCGCGCCAAAAAAGGATTAAGAGAGCGCAGCACCAAAAAACCTTGATGAGAGT	4057
Qy	1301	ThrProIySGLuGIuProThMetIySThrSerIySGLySerIySLeuIySGLySGLySGLyS	1320
Dd	4058	ACACCAAAAGAGAGCCCAACAATGAATAACTTCTAAAAAAGGAGACCAAAAAAATCTGGA	4117

QY	1331	ProProAlaAProGluInGluInGluInGluInuArgInSerGlyAsnThrGluInuLys	1340
DB	4118	CTCTCAGACACAGAGAGAGAGAAAGAAAGACAAAGTGGAAATACGGAACAGAG	4177
QY	1341	SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu	1366
DB	4178	TCCAAAAGCAAAACAGCACCAGATGTGTCAAGAGAGACAGACAGAGACGAATCTCTGAA	4237
QY	1361	SerSerAlaAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys	1380
DB	4238	TCTAGTGTCAATGTGATCCACACAGTCCACACACAGAAAGAGACGAGAAACATCAAAA	4297
QY	1381	ThrProSerProSerGlnProLysLysAsnVal	1391
DB	4298	ACGCCATACCATCACAACCAAAAAAATGTC	4330
RESULT 4			
ID	ABV28244		
XX	ABV28244	standard; cDNA; 7473 BP.	
XX	ABV28244;		
AC	16-SEP-2002	(first entry)	
DT			
XX			
DE		Human prostate expression marker cDNA 28235.	
XX			
KM		Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	
KW		pharmacogenomic marker; gene; ss.	
XX			
OS		Homo sapiens.	
XX			
PN	WO200160860-A2.		
XX			
PD	23-AUG-2001.		
XX			
PF	20-FEB-2001; 2001WO-US05171.		
XX			
PR	17-FEB-2000; 2000US-183319P.		
PR	16-MAR-2000; 2000US-189862P.		
PR	25-MAY-2000; 2000US-207454P.		
PR	09-JUN-2000; 2000US-211314P.		
PR	18-JUL-2000; 2000US-219007P.		
PR	13-DEC-2000; 2000US-255281P.		
XX			
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
XX			
PI	Schlegel R, Endege WO, Monahan JE;		
XX			
DR	WPI; 2001-662795/76.		
XX			
XX		Novel isolated nucleic acid molecule associated with cancerous state of	
PT		prostate cells and correlating with presence of prostate cancer, useful	
PT		for detecting presence of prostate cancer, stage of prostate cancer -	
PS	Claim 1; Page 5869-5870; 11750pp; English.		
XX			
CC		The invention relates to an isolated nucleic acid molecule (I) comprising	
CC		a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the	
CC		specification or its complement. (I) is useful for:	
CC		(a) assessing whether a patient is afflicted with prostate cancer;	
CC		(b) monitoring the progression of prostate cancer in a patient;	
CC		(c) assessing the efficacy of a test compound to inhibit prostate	
CC		cancer in a patient;	
CC		(d) assessing the efficacy of a therapy for inhibiting prostate cancer	
CC		in a patient;	
CC		(e) selecting a composition for inhibiting prostate cancer in a patient;	
CC		(f) assessing the prostate cell carcinogenic potential of a compound;	
CC		(g) determining whether prostate cancer has metastasized in a patient;	
CC		(h) assessing the aggressiveness or indolence of prostate cancer in a	
CC		patient;	
XX		(I) is also useful as a pharmacodynamic or pharmacogenomic marker.	
XX			

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1: Page 5869-5870; 11750bp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

- (a) assessing whether a patient is afflicted with prostate cancer;
- (b) monitoring the progression of prostate cancer in a patient;
- (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
- (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
- (e) selecting a composition for inhibiting prostate cancer in a patient;
- (f) assessing the prostate cell carcinogenic potential of a compound;
- (g) determining whether prostate cancer has metastasized in a patient;
- (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Claim 1; Page 5869-5870; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

- (a) assessing whether a patient is afflicted with prostate cancer
- (b) monitoring the progression of prostate cancer in a patient;
- (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate cell carcinogenic potential of a compound

(g) determining whether prostate cancer has metastasized in a patient

(ii) assessing the aggressiveness of involvement of prostate cancer in patient;

(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 7473 BP: 2458 A: 1244 C: 1479 G: 2285 T: 7 other:

Alignment Scores:

Pred. No.:	0	Length:	7473
Score:	7149.00	Matches:	1386
Percent Similarity:	99.64%	Conservative:	0
Best Local Similarity:	99.64%	Mismatches:	5
Query Match:	99.39%	Indels:	0
DB:	23	Gaps:	0

US-09-512-581b-2 (1-1391) x ABV28244 (1-7473)

QY 1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProGlyValIlys 20
DB 158 ATGGCTATTCAAAGACTAGGACCAATGATGAAAAATTCATATCCGCTGGGGTCAAG 217
QY 21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValIlys 40
DB 218 GAATATACATATAAATATCTAAAGAGAGATGCTGAGACGATTAAGATGCTGTGAAA 277
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
DB 278 ACTTTATGGATATGACAGAGACTCTGAGAGAAAAAGAGCTTTATTAACCTTAGCT 337
QY 61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
DB 338 TTACATCTGCTTACAGATTTTCTTCACAGATCTCTATAAGATGTTCCGTTACTGTA 397
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
DB 398 GCCCTGCTGCTGCTGATATTTTCAGATTTATCTCTCGAAGCTCTTACACATCCCT 457
QY 101 AspLysLeuLysAspIlePheMetPheIlePheArgGlnLeuLysGlyLeuGluAspThr 120
DB 458 GATTAACATAAAGGATATATTTATGTTATTAACAAGAGAGGAGGCTAGAGGATACA 517
QY 121 LysSerProGlnPheAsnArgTyrPheTyrLeuGluLysAsnIleAlaIlePheValLysSer 140
DB 518 AAGGCCCAACATTCATACCTATTTTATTTACTTGAGAACATGCTTGGGTCAAGTCA 577
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
DB 578 TATTAACATATGCTTTGAGTTAGAAGATAGCAATGAATTTTACCCAGCTATACAGAAC 637
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 638 TTATTTTCACTTATTAACAATGGCCACAATCAGAAAGTCCATATGACATGGTAGACCTT 697
QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
DB 698 ATGAGCTCATTTATTTGTGAAGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTAA 757
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
DB 758 GTAATATGCTTACCTGCTCATTAAGAAATTTAAACAAGACATATGATTTGGCAAAAGCT 817
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIlePheThrPhePheAsnGlnVal 240
DB 818 TTACTGAGAGGAGAGCTCAAGCTATTTAGCCATATATACCAATTTTATTAATCAGAGTT 877
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 878 CTGATGCTTGGAAAAACATCTATCAGCGATTTGTCAAGCATGCTTTGACTTAATTTTG 937
QY 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 938 GACCTCTACAAATATGATGATGCTATTTGCTCTCTCTTTTACCCTGATTTAA 997
QY 281 LeuLysSerAsnAspAsnGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
DB 998 TTAAAGAGCAATGATATGAGAGAGCCCTTCAAGCTTTTAAACTACTGCAAAAATGTTT 1057
QY 301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrGlnCysTyrLeuGly 320

DB 1058 GGGCAAAAGATTCAGATTCGCTTCATAAACAAGCACCTTGGAGTGTACTTGGCC 1117
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
DB 1118 AGGTTTATGATATTCATTCATGATCCAGTCCCTCGAAGTGTGAATTTGCTAGCCATTTGT 1177
QY 341 LeuMetAsnHisProAlaPheAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
DB 1178 CTCATGAACCAATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAGTAGGTCACAT 1237
QY 361 AspProGluGlnAlaIleArgHisAspValIleValSerIleValIlePheAlaLysLys 380
DB 1238 GACCTCAGAGAGCTATTAGACATGATTTTGTGTCANATGATACAGCTGCTAAAAAG 1297
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
DB 1298 GATATCTCTGCTGCTCAATCATCACTTAAATTTGTGAGAGAGAACATTTAGACAAA 1357
QY 401 ArgThrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysTyrAla 420
DB 1358 CGATGAGAGAGTACCAAAAGAGCCATGATGGAGCTTCCCAAAATTTATTAAGAAATATGCT 1417
QY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaIlePheLysAspLysLeu 440
DB 1418 TTACAGTCAGAGCTGGAAGAAAGATGCTGCAAAACAGATGATGATCAAGACAAATTTG 1477
QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
DB 1478 CTACATATATATATTCAAAATATGATGATGATGATGATGATGATGATGATGATGATG 1537
QY 461 GluTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
DB 1538 CAATACATAGGTTCTCTCAAAATTTAGAACTACAGACAGAGGAATGCTTATATTTACTTG 1597
QY 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetCtyrPlysCysGlnAsn 500
DB 1598 TATCCACACAGCTGATTTAAATGCTGTGAAGACATTTGAATGAATGTGCAAAAT 1657
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
DB 1658 CTGCTCCGACATCAAGTAAAGATTTGCTTGCATTTGATTAAGCAACCAAAACAGATGCC 1717
QY 521 SerValLysAlaIlePheSerLysValMetValIlePheArgAsnLeuProAspProGly 540
DB 1718 AGTGTCAAGGCCATATTTTCAAAAGTATGCTTATTAACAAGAAATTTTACCTGATCCGTG 1777
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
DB 1778 AAGGCTCAGAGATTTCAAGAAGAAATTCACACAGGTTGTAGAAATGATGAGAAATTAAGA 1837
QY 561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
DB 1838 AAGCAGTTAGAAAGTACTTGTATGCCAACATGCTCCGCAAGCAGGCTGAAGGTTGTGTG 1897
QY 581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
DB 1898 CGTAAATTAACATAAGAGTTGGGCAACCCCAACAGCTTACAAATCTTTCTCTGGAATG 1957
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
DB 1958 ATCAAGTTTCTCTTGGAGAGATAGCACCTGTGCACATATGATACCAATCTATCAGTGTCT 2017
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
DB 2018 CTTATTTAAACAAGTAAATCAATATGATGGAACACACATGATGTAAGATGAGAGGTGT 2077
QY 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValIleLeuSerPheThrHis 660
DB 2078 CCAACTATTAAGCCATCAAGCAGAGGTTCTTGAAGCTTTAAGGACTCTCATTTACAT 2137
QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680

Db 2138 CCCATCTCATTTTCCTGCTGTAACATTTGAACTATTCAGCTTGTCTGAAATATGAT 2197
QY ASPGLuLySValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2198 GATGAAAAAGTAGACAGACTGCATACAAATTTTCAAAAACACAGAACCAAAATTTGAA 2257
QY 701 GLuAspPheProHisIleArgSerAlaLeuProValIleuHisIleLysSerLys 720
Db 2258 GAGGATTTTCCACACATCAGATCAGCTTGTCTTCCGTTTACATCACAATTTAAAAA 2317
QY 722 GLYProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLys 740
Db 2318 GACCCCCCGCTAACCCAAATATGCAATTCATTCATTCATGCAATTTTCTAGTAAA 2377
QY 741 GLuThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2378 GAGACCAGATTGTCACAGATATTGAGCCCTGCTGATTAAGAGCCTACATCCAAACCTG 2437
QY 761 GLuHisLeuIleThrProLeuValIleIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2438 GAACATCTCATTAACACATTTGGTTACTATTGGTCATATTGCTCTCTGCACTGATCAA 2497
QY 781 PheAlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2498 TTTGCTGCTCTTGGAAATCTTTGGTAGCTATTGATGTAAGAATCTTCATGAAAT 2557
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
Db 2558 GATCGGCTTCCAGGAAAAAGACAACTTAACCTTGGGTTCCAGATGAGAAGATCTCCT 2617
QY 821 GLuThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
Db 2618 GAACAAATGGTCAAAATTCAGGCTATTAAATGATGGTTCGATGGCTACTTGGAAATGAA 2677
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
Db 2678 AATTAATCACAGTAATAACAGGAAGTTCTACTTAAGTTGCTAACAAATATTCATAGT 2737
QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2738 GATGGAGACTGTGACAAACAGGGGAAATTAAGTAAACAGATATGTGCACGCTGAGACTT 2797
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2798 GCTGCTGGAGTCTATGTGGAAGCTGGCAAGAACCCGTTACCATGTAATCATCACA 2857
QY 901 LeuGlnGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2858 TTGGAACAATATCAGCTATGTCATTAAGCATATCAGATGAATATGCTATCAAGTAAGCAA 2917
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Db 2918 GTGTGTGGCCAGAAACTTCACAAAGGCTTCCCGTTTACGGCTTCCACTTGAATATAG 2977
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Db 2978 GCATATGTGCTTGTGCAAAAAGATCCGTAAAGAGAGAGAGCTCATGCTAGGCAA 3037
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisIleAlaValSer 980
Db 3038 TGTGTGTGAAAAATTAATATTAAGCGGAGATATCTGAAAGCAGCTGAGCTGTAGT 3097
QY 981 GLuLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3098 GAAAAATTAATGTCTCTTACAGAGATATGTTGCCATATACAAATTCACCTTTGGCA 3157
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3158 CATGACCAGATATATGCAAAAGTACAGAGATATGCAACAACTTAAGATGTTAAAGATGT 3217
QY 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3218 CTTTGGTTGTCTGGAATATTAATGCTAAAAATGAATAACAGTCAAGCTTTTATC 3277

QY 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3278 AGAAGATGTGAAAAATATTAACAAACAAAAAGATGCCCAAGACCAATGATGCAAAA 3337
QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3338 ATGAATGAATAAACTGTACACTGTGTGATGTTGTCATGAATATCATATGTCAAAGACT 3397
QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaTrpPheThrGln 1100
Db 3398 ACTACATACATTTGGAAATCTCTTAAGACCCGGTACTCTCCAGCTGTTTCTTCACTCA 3457
QY 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGluLysSerPhePhe 1120
Db 3458 CCGACAAAGAAATTTAGTAACACCAAAATTAATCTCCTCGTAAGTAAGTAATCATTTTTC 3517
QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3518 ACTCTGTGAAAAACCTTAACCAACCAATGTTTCAAGAGCTGTAAACAAAGCCACTTTCATCA 3577
QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3578 GCAGGCAAGCAATCTCAGACCAATATCATACGATGGAACCTTAAAGCATGCAAGCAC 3637
QY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3638 AGCTCAATTCCAAGCTCTCCTGGAAGATTAAGAGAGAGGAGCTGATGATCTGCAATAGAT 3697
QY 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3698 CACAGTGAATAAATTAACAAATGCTCTCACCTTTCCTCCGGGAAAAAAGTGCACAAG 3757
QY 1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
Db 3758 AGAGAGACTCTGATCTTGTAAAGTCTGAATTTGAGAGAGCTTGAAGCAGGAAAAAAG 3817
QY 1221 ProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
Db 3818 CCCGTACAGAACAGAGAGGAAATTAAGTATGATGACTGACTAAAGTTGGTACAGGAA 3877
QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgIleHisThrAlaSerGluSerAsp 1260
Db 3878 CACAAACCTTAAGAGCGCTGACGCAAGTCGGAAGAGGCCATACGCGCTTCAGAAATCTGAT 3937
QY 1261 GlnGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluLysAspGlu 1280
Db 3938 GAACACAGTGGCCTGAGGAAAAAGAGCTCAAAAGAAATATATTAGAAATGAAGATGAA 3997
QY 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
Db 3998 CAGAAATAGTCCGCCAAAAAAGGGTAAAGGCGCCGACCAACCAAAACCTTGTGTGAGGT 4057
QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
Db 4058 ACACCAAAAAAGAACCAACAAATGAATAAATCTTAAAAAAAGAAACAAAAAATATGGA 4117
QY 1321 ProProAlaProGluGluGlnGluGlnGluGlnGlnGlnSerGlyAsnThrGlnGlnLys 1340
Db 4118 CTTCCAGCAACCAAGAGAGGAGGAAGAAGAAAGAAACAAAGTGAATATGCGAAGCAAG 4177
QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4178 TCCAAAAAGCAAAACAGACGAGTCAAGAGAGCAGACAGACAGATCTCTCGAA 4237
QY 1361 SerSerAlaIleLeuSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4238 TCTAGTGCAATTAATTCACACAGTCCACCAACCAAAAGGAGAGAGACCATCAAAA 4297
QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4298 ACGGCATCACCATCAACCAAAAAAATATGTC 4330

RESULT 5
 ABV25469
 ID ABV25469 standard; cDNA; 7473 BP.
 AC ABV25469;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 25460.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 5055-5057; 11750pp; English.
 CC
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement, (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 CC
 CC
 SQ Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;
 XX
 Alignment Scores:
 Pred. No.: 0 Length: 7473
 Score: 7148.00 Matches: 1386
 Percent Similarity: 99.64% Conservative: 0
 Best Local Similarity: 99.64% Mismatches: 5
 Query Match: 99.37% Indels: 0
 DB: 23 Gaps: 0
 US-09-512-581B-2 (1-1391) x ABV25469 (1-7473)
 OY 1 MetAlaHisSerIyThArGthraSnaSpGlyLysIleThTyRPrProGlyValLys 20
 DB 158 ATGGCTCATTCGAAACACTRGGCCCAATGATGCAAAAATTACATATCCGCTGGGGTCAG 217
 OY 21 GtulleSraSpLysIleSerIySgluGluMetValArgArgLeuIlyMetValLys 40

Db	218	GAATAATCTCGAATAAATATCTAAAGGAGATGGTGAACGATTAAAGATGGTGTGAAA	277
QY	41	ThrPheMetAspMetaspGlnAspSerGluGluGluLeuLeuLeuValAla	60
Db	278	ACTTTATATGATATGACACGACGCTCTGAAGAAGAAAGAGACTTTATTTAAACCTAGCT	337
QY	61	LeuHISLeuAlaSerAspPhePheLeuLysHISProGluLysAspValArgLeuLeuVal	80
Db	338	TTACATCTTGCTTCAGATTTTTTTTCTCAAGCATCCCTGATAAAGATCTTGCGCTTACTGGTA	397
QY	81	AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro	100
Db	398	GGCTCTCGCTTCGATATTTTTCAGAGATTATGCTCTGAAACCTCTTACACATCCCT	457
QY	101	AspLysLeuLysAspIlePheMetPheIleThrArgGluLeuLysGluLeuAspThr	120
Db	458	GATAAACAATAAGCATATATTTATGTTTATTAACACAGACAGTTGAAAGGGCTACAGGATACA	517
QY	121	LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaIleArgValLysSer	140
Db	518	AAGACCCCAACATTCATAGATATTTTATTAACCTGAGAACACTTGCTGGGCAAGTCA	577
QY	141	TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGluLeuTyrArgThr	160
Db	578	TATACATATGCTTGAGTTAGAACAGATACCAATGAATTTTCCACCAGCATACACAAACC	637
QY	161	LeuPheSerValIleAsnAsnGluHISAsnGluLysValHISMetHISMetValAspLeu	180
Db	638	TTATTTTCAGTTATAAACAATGGCCCAACATCGAAAGTCCATATGGACATGGTAGACCTT	697
QY	181	MetSerSerIleIleCysGluGluLysAspThrValSerGlnGluLeuLeuAspThrValLeu	200
Db	698	ATGACCTCATATTTTGTGAAGGTATACAGTGTCACAGAGCTTTTGGATACGGTTTTA	757
QY	201	ValAsnLeuValProAlaHISLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla	220
Db	758	GTAATCTCGTGACCTGCCTCATGAATTTAAACACAGCAACATATGTTGGCAAGGCT	817
QY	221	LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGluVal	240
Db	818	TTACGGAAGAGGACAGCTTAACCTATGTAGCCATATATTAACCAATTTTTTTTAAATCAGGT	877
QY	241	LeuMetLeuGluLysThrSerIleSerAspLeuSerGlnHISValPheAspLeuIleLeu	260
Db	878	CTGATGCTTGGGAAAAACATCTATCAGCCATTTGTGCAGACATGTCTTGACTTAATTTTG	937
QY	261	GluLeuTyrAsnIleAspSerHISLeuLeuLeuSerValLeuProGluLeuGluPheLys	280
Db	938	GAGCTCTCAAAATATGATAGTATTTGGCTCTGCTCTGTTTACCCACGCTTGAATTTAAA	997
QY	281	LeuLysSerAsnAspAsnGluGluLysArgLeuGluValValLysLeuLeuAlaLysMetPhe	300
Db	998	TTAAAGACCAAGATAATGTAGAGGCGCTACACAGTTGTTAACTACTGGCAAAAAAGTTT	105
QY	301	GluAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrGlnGlyTyrLeuGly	320
Db	1058	GGGGCAAAAGATTACGAATTTGGCTTCTCAAAACAAAGCCACTTTGGCAGTGCTACTGGGG	111
QY	321	ArgPheAsnAspIleHISValProIleArgLeuGluCysValLysPheAlaSerHISCys	340
Db	1118	AGGTTATATGATATTCATGCTACCAATCCGCTGGAATGTGTGAAATTTTCTGAGCCATTTG	117
QY	341	LeuMetAsnHISProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHIS	360
Db	1178	CTCAATGAACCATCTGATTTTACCAAAAGCTTTAACAGATATCTTAAAGTGAAGTACAT	123
QY	361	AspProGluGluAlaIleArgHISAspValIleValSerIleValThrAlaAlaLysLys	380
Db	1238	GACCTGAGGAAGCTATTAGACATGATGTATTGTGTCTCAATAGTTACAGCTGCTAAAG	129
QY	381	AspIleLeuLeuValAsnAspHISLeuLeuAsnPheValArgGluArgThrLeuAspLys	400

Db 1298 GATATTCTTGGTCATGATCATCTTAAATTTTGTGAGAGAGAACATTAGACAAA 1357
QY ArgTPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
Db 1358 CAGTGGAGAGTACGCAAAAGAGCCATGATGGGACTTCCCAATTTATTAAGAAATATGCT 1417
QY LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTPrIleLysAspLysLeu 440
Db 1418 TTACAGTCAAGAGCTGGAAAGATGCTGCAAAAACAATAGCATGATCAACAAATG 1477
QY LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuValGluIlePheAla 460
Db 1478 CTACATATATATTAATAATAGTATGTGATGATCGACTGTGTGAACGGATCTTGTCT 1537
QY GlnTyrMetValProHisAsnLeuGluThrThrGlnArgMetLysCysLeuTyrTyrLeu 480
Db 1538 CATACATATGGTCTCACAATTTAGAACTACAGAAAGGATGAAATGCTTATTTACTTC 1597
QY TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTPrLysCysGlnAsn 500
Db 1598 TATGCCACACTGGATTAAATGCTGTGAAAGCATGTGAATGAATGTGAAATGTCAAAAT 1657
QY LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
Db 1658 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTGATTAGCAACCCCAACAGATGCC 1717
QY SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
Db 1718 AGTGTCAAGGCCATATTTTCAAAAGATGGTATTACAAAGAAATTTACTGTACTCTGCT 1777
QY LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
Db 1778 AAGGCTCAGAGATTTCATGAAGAAATTCACACAGCTGTAGGATGATGAGAAATAGCA 1837
QY LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnIleAlaGluLysVal 580
Db 1838 AAGCACTTGAAGACTTGTGTACTCCAACTGCTCTGTCAGAGAGGCTGAAGGTGTG 1897
QY ArgGluIleThrLysLysLeuGluAsnProLysGlnProThrAsnProPheLeuGluMet 600
Db 1898 CGTGAATATACATAAGAGTTGGCCAAACCCCAACAGCTTCAAAATCTTCCCGGAATG 1957
QY IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
Db 1958 ATCAAACTTCTCTGTGAGAGGATAGCACCTGTGCATAGATACCAATCTATCACTGCT 2017
QY LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal 640
Db 2018 CTTATTAAACAAAGTGAACAATCAATAGATGAGAACAGCATGATGATAAGTGAAGGTGT 2077
QY ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
Db 2078 CCAAGTATCAACAGCCATCAGAGCAGGCTTGAAGCTGTTAAGTACTCTCATTTACACAT 2137
QY ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Db 2138 CCCATCTCATTTCTATCTGCTGAACAATTTGAATCATTTACTGCTTGTGAAAAGGAT 2197
QY AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2198 GATGAAGAAAGTACACAGAGCTGCATCAAAATTTTCAAAAACAGAGAACCAAAATTTGAA 2257
QY GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLysLys 720
Db 2258 GAGAGATTTTCCACACATCAGATCAGGCTTCCGTTTACATCACAAATGTAAAAAA 2317
QY GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLysLys 740
Db 2318 GGGACCCCGCCGTCACCAAAATATGCCATTTGATTCATTCATCGATATTTTCTAGAAA 2377
QY GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2378 GAGACCCAGTTTGCACAGATATTGTAGACCTCTGCATAGAGCCTAGATCCAAACCAACTG 2437

QY GluHisIleLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2438 GAACATCTATAACACCATTTGGTACTATTTGGCATATTTGCTCTCTCCACTGATCA 2497
QY PheAlaAlaProTPrLysSerTPrValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2498 TTTGGCTGCTCTTTGAAATCTTTGGTACTTTTCATTTGTGAAGATCTTCTCATGAT 2557
QY AspArgLeuProGlyLysLysThrThrLysIleTPrValProAspGluGluValSerPro 820
Db 2558 GATGGCTTCCAGGGAAGAAAGACACTAAACTTGGGTTCCAGATGAAGATATCTCT 2617
QY GluThrMetValLysIleGlnAlaIleLysMetMetValArgTPrLeuGluMetLys 840
Db 2618 GAGACAAATGGTCAAAATTCAGGCTATTTAAATGATGGTTGATGCTTGGAAATGAAA 2677
QY AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThrIleLeuHisSer 860
Db 2678 AATTAATCACAGTAAATCACAGAACTTCTACTTAAGATTTGCTTAACAACAATATTTGCATAGT 2737
QY AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2738 GATGGAGACTTGACAGAACAGGGGAAATTAGTAAACCAAGATATGTACAGCTGAGACTT 2797
QY AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2798 GCTGCTGGGAGTGTCTATTGTGAAGCTGGCAACAAGAACCTGTACCATGAAATCATACAA 2857
QY LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2858 TTGAACAATATACGATATGTGCTATGCTATCAACAGATGCAATACAGTAAAGACAA 2917
QY ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Db 2918 GTTGTGGCCAGAACTTCAACAAAGSCCTTTCCGTTTACGGCTTCCACTTGAGTATAG 2977
QY AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Db 2978 GCATCTGTGGCCCTTTGTGCAAAAAGATCCGTAAAGAGAGAAAGACTCATGCTAGGCAA 3037
QY CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaLysSer 980
Db 3038 TGTTTGTGTAATAATTAATGAAGCGGGAATATCTGAAGCAGCATCCAGCTGTTAAT 3097
QY GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3098 GAATAATTTATTTGCTCTTCTTACCAGAGTATGTTGTTCCATATACATTTACCTTTGGCA 3157
QY HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3158 CATGACCAGGATATATCAAAAGTACAGATATTGAACAACCTTAAGATGTAAAGAAATGT 3217
QY LeuTPrPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3218 CTTTGTGTTGTTCTGGAATATTTAATGCGCTAAAGAAATGAAGATACAGTACCGCTTTATC 3277
QY ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3278 AGAAAGATGTGAAATAATTTAAACAACAAAGATGCCCAAGAGCCAGATGATGCAAAA 3337
QY MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3338 ATGAATGAAGAAACTGTACACTGTGTGATGTTGCCATGAATCATCATATGTCAAAGAGT 3397
QY ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
Db 3398 ACTACATACAGTTTGAATCTCTTAAGACCGGTACTACACAGCTGTTTCTTCACTGCA 3457
QY ProAspLysAsnProPheSerAsnThrLysAsnTyrLeuProGluLysSerPhePhe 1120
Db 3458 CCTGACACAGATTTTCAGTAAACCAAAATTAATCTCTCTCTGTAATAATGAATCATTTTTC 3517

QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
 |||
 Db 3518 ACTCCGGAAAAACCTTAACCAACCATGTTCTTAGAGCGCTTACAAAGCCTTTATATCA 3577
 |||
 QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
 |||
 Db 3578 GCAGCGCAACCAATCTCAGACCAATCATCAGCAATGAAAGCTGTAAGCAATGCAACGACG 3637
 |||
 QY 1161 SerSerAsnProSerSerProGlyArgGlyLeuLysGlyArgLeuAspSerSerGluMetAsp 1180
 |||
 Db 3638 ACCTCAATCTCAAGCTCTCTCGAAGAAATTAAGGGAGCGCTGATGTTCTGAAATGAT 3697
 |||
 QY 1181 HisSerGluAsnGlnLysPylThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
 |||
 Db 3698 CACAGCGAAATGAAATTAACATATGTTTCACTTTCCCGGGGAAAAAAGTGACAAAG 3757
 |||
 QY 1201 ArgAspAspSerSerAspLeuValArgSerGluLeuGlyLysProArgLysArgLysLysThr 1220
 |||
 Db 3758 AGAGACGACTGTGATCTTCTTAAGGTCTGAATTTGAGAACCTTAAGACGAGGAAAAAAGC 3817
 |||
 QY 1221 ProValThrGluGlnGluGlnLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
 |||
 Db 3818 CCGCTGCAGAACAGAGAGAGAAATTAAGTATGATGACTTAAAGTTGGTACAGGAA 3877
 |||
 QY 1241 GlnLysProLysGlySerGlnArgSerArgLysArgLysHisThrAlaSerGlnSerAsp 1260
 |||
 Db 3878 CAGAAACCTAAAGGACAGTACGAGAGTCGGAAGGCAATACGCTTCAAGATCTGAT 3937
 |||
 QY 1261 GlnGlnGlnThrProGlnGlnLysArgLeuLysGlnAspIleLeuGlnLysAspGlu 1280
 |||
 Db 3938 GACACGACAGTGGCTGAGAGAAAAAGCGCTCAAGAAATATTTAGAAATGAAGATGAA 3997
 |||
 QY 1281 GlnAsnSerProProLysLysArgGlyArgGlyArgProProLysProLeuGlyGlyGly 1300
 |||
 Db 3998 CAGATATGTCGCCCAAAAAAGGCTTAAGAGGCCGACCAAACTCTTGTTGGAGGT 4057
 |||
 QY 1301 ThrProLysGlnGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
 |||
 Db 4058 ACACCAAAAAAGAGACCAACATGAACCTTAAAAAAAGAAAGCAAAAAAAATCTGGA 4117
 |||
 QY 1321 ProProAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1340
 |||
 Db 4118 CCTTCAGACACAG 4177
 |||
 QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlu 1360
 |||
 Db 4178 TCCAAAAGCAAAACACACCGAGCTGTCAAGAGAGACACACAGAGAGCAAAATCTCTGAA 4237
 |||
 QY 1361 SerSerAlaIleGlnSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
 |||
 Db 4238 TCTACTGCATTTGATTCACACAGTCCACACCAAGAAAGGAGGAAAGACCATTCAAAA 4297
 |||
 QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
 |||
 Db 4298 ACGGCATCATCCATCACACCAAAAAAAATATGTC 4330
 |||
 RESULT 6
 ABX14056
 ID ABX14056 standard; cDNA; 6744 BP.
 XX
 AC ABX14056;
 XX
 DT 24-FEB-2003 (first entry)
 XX
 DE cDNA encoding human SCC-112 tumour suppressor gene.
 XX
 ss; gene; human; chromosome 4p14; SCC-112; cancer; apoptosis; alldodynia;
 KW degenerative disorder; metastasis inhibition; breast cancer; causalgia;
 KW kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
 KW squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
 KW global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
 KW hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;
 KW cardiac arrest; spinal cord lesion; stomach cancer; lung cancer;

KW neonatal distress; Alzheimer's disease; uterine cancer; colon cancer;
 KW multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
 KW Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
 KW amyotrophic lateral sclerosis; cell survival; cell proliferation;
 KW tumour suppressor.
 KW
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 232..4125
 FT /tag= a
 FT /product= "Human SCC-112"
 FT polyA_signal 6721..6726
 FT /tag= b
 FT /standard_name= "PolyA signal"
 FT
 XX WO200281641-A2.
 PN
 PD 17-OCT-2002.
 XX
 XX 08-APR-2002; 2002MO-US10850.
 PF
 XX 06-APR-2001; 2001US-281780P.
 PR
 XX (GEOU) UNIV GEORGETOWN.
 PA
 XX Kasid UN, Kumar D, Ahmad I;
 PI
 XX WPI; 2003-103330/09.
 DR
 XX P-PSDB; ABG72803.
 XX
 PT New isolated SSC (undefined) tumor suppressor polypeptides and
 PT polynucleotides, useful for diagnosing, preventing or treating cancer
 PT or degenerative disease, e.g. Alzheimer's Disease, Huntington's
 PT disease, or multiple sclerosis
 PT
 PS
 XX Claim 1: Fig 1B; 83bp; English.
 XX
 CC The invention relates to a new isolated polypeptide SCC-112. The SCC-112
 CC polypeptides and polynucleotides are useful for diagnosing, preventing or
 CC treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
 CC pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
 CC cancer, stomach cancer, colon cancer, lung cancer); or degenerative
 CC disease or disorder (e.g. global and focal ischaemic and haemorrhagic
 CC stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
 CC damage, nerve cell damage caused by cardiac arrest or neonatal distress,
 CC epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
 CC pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
 CC hyperalgesia, alldodynia, Alzheimer's disease, Huntington's disease,
 CC Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
 CC Korsakoff's syndrome). The polypeptides and polynucleotides are also
 CC useful for inducing apoptosis in cancer cells, increasing survival or
 CC proliferation of a cell, or inhibiting cancer cell proliferation and/or
 CC metastasis in a cancer patient. The polynucleotides can be used as probes
 CC to detect complementary nucleotide sequences, or as primers to obtain
 CC additional copies of the polynucleotides. SCC-112 may also be used for
 CC identifying drugs for treatment of cancers. The present sequence
 CC represents cDNA encoding for the human SCC-112 tumour suppressor gene which
 CC is located on chromosome 4p14.
 CC
 XX Sequence 6744 BP; 2156 A; 1243 C; 1400 G; 1945 T; 0 other;
 SQ
 XX
 Alignment Scores:
 Pred. No.: 0
 Score: 4760.50 Length: 6744
 Matches: 935
 Percent Similarity: 80.29% Conservative: 173
 Best Local Similarity: 67.75% Mismatches: 210
 Query Match: 66.18% Indels: 63
 DB: 25 Gaps: 15
 US-09-512-581B-2 (1-1391) x ABX14056 (1-6744)
 QY 7 ArgThrAsnAspGlyLysIleThrThrProProGlyValLysGlnIleSerAspLysIle 26

Db 160 CGCACACACGCCGGAAGATCGCTTACCTCCGGGGGTAAAGAGATCACCGACAAATC 219
Qy 27 SerLysGIuLumetValArgArgLeuLysMetValValLysThrPheMetAspMetAsp 46
Db 220 ACCACGACGACGATGATCAAAACGCTCAGATGGTATGAAACCTTTATGATATGAT 279
Qy 47 GluAspSerGIuGIuLysGIuLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db 280 CAGAGCTCAGAAAGATGAAAAACAGCATATCTCCACTAGCCTTGACATCTTGATCGAA 339
Qy 67 PhePheLeuLysHisProGIuLysAspValArgLeuLeuValAlaCysCysLeuAlaAsp 86
Db 340 TTCTCTCCAGGAACCCCAATAAAGATGGCTGCTCTTGATGATGTGGTGGCAT 399
Qy 87 IlePheArgIleTyrAlaProGIuAlaProTyrThrSerProAspLysLeuLysAspIle 106
Db 400 ATCTTGCTGATCTATGCCCGCAAGAGCTCATATCTTCCCATGTAAACCTTAAGACATA 459
Qy 107 PheMetPheIleThrArgGIuLeuLysGIuLeuGIuAspThrLysSerProGIuPheAsn 126
Db 460 TTTTGTGTTATTACACACATTTAAAGTTTGAGATACAAAGAGTCCACAGCTTTAAT 519
Qy 127 ArgTyrPheTyrLeuLeuGIuAsnIleAlaTyrValLysSerTyrAsnIleCysPheGIu 146
Db 520 AGAATCTTTATTATTAGAAATTTAGCTTGGGTAAATCATATTAACATCTGCTTGAA 579
Qy 147 LeuGIuAspSerAsnGIuIlePheThrGIuLeuTyrArgThrIlePheSerValIleAsn 166
Db 580 TTGGAAATTCGAATGAATTTTATTCAGCTTTTGAAGACTCTCTCTCAGTATCAAC 639
Qy 167 AsnGIuHisAsnGIuLysValHisMetHisMetValAspLeuMetSerSerIleLeCys 186
Db 640 AATGACCATATTAAGAGTACAAATGACATGCTAATTTGATGATGATTCATCATCATG 699
Qy 187 GIuGIuLysPheThrValSerGIuGIuLeuLeuAspPheThrValLeuValAsnLeuValProAla 206
Db 700 GAAGGTATGAGTACTCAAGAAATTTATGGGCTCCATCTTATTAAACCTCATTCCTGCA 759
Qy 207 HisLysAsnLeuAsnLysGIuAlaTyrAspLeuValAlaLysAlaLeuLysArgThrAla 226
Db 760 CATAGAACTTAATATAACAGTCTTGACCTTGCAAAAGTGCATTTGAAAGAAAGAGTGC 819
Qy 227 GluAlaIleGIuProTyrIleThrThrPhePheAsnGIuValLeuMetLeuGIuLysThr 246
Db 820 CAGACTATGAGGATCATGCTCCAAATTTTTCATTAATCAAGCTCGGCGGGAATCA 879
Qy 247 SerIleSerAspLeuSerGIuHisValPheAspLeuIleLeuGIuLysTyrAsnIleAsp 266
Db 880 TCAGTATGATATTTGTGAGAACATGTATTTGATCTGATTCAGGAATTTTGTCTATGAT 939
Qy 267 SerHisLeuLeuLeuSerValLeuProGIuLeuGIuPheLysLeuLysSerAsnAspAsn 286
Db 940 CCGCATTTATTATATCCGTCATGCCACAGCTGTAATCAAACTTAAGAGAGATGATGGA 999
Qy 287 GIuGIuLysArgLeuGIuValValLysLeuLeuAlaLysMetPheGIuAlaLysAspSerGIu 306
Db 1000 GAAGAGGATTAAGCTGTGGTTCGACTTCTAGCTAAATTTGTTGGCTCCAAAGATTCGAT 1059
Qy 307 LeuAlaSerGIuAsnLysProLeuTyrProGIuCysTyrLeuGIuLysArgPheAsnAspIleHis 326
Db 1060 TTGGCACACAGAAATCGTCTCTTTGGCAATGTTTCTTGACATTTAATGATATTCAT 1119
Qy 327 ValProIleArgLeuGIuLysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
Db 1120 GTTCCGTGATGATAGAAAGTGTAATTTGGCCAGTCAATGTTTAATGAATCACCCAGAT 1179
Qy 347 LeuAlaLysAspLeuThrGIuTyrLeuLysValArgSerHisAspProGIuAlaIle 366
Db 1180 TTGACGAGAGATCTCACAGATATTTAAAGTTGATGATCACATGATCCAGAAAGACTATT 1239
Qy 367 ArgHisAspValIleValSerIleValThrAlaAlaLysAspIleLeuLeuValAsn 386

Db 1240 CGTCATGATGTCATTTTACTATATATAACAGCTGCCAAGAGGACCTGGCCTTAGTAAT 1299
Qy 387 AspHisLeuLeuAsnPheValArgGIuArgThrLeuAspLysArgTyrPheValArgLys 406
Db 1300 GATCAGCTGCTTGCTTGTATAGGGAAGAAACACTGATATTAACCGTGGCCAGTAAAGAAA 1359
Qy 407 GIuAlaMetMetGIuLeuAlaGIuIleTyrLysLysTyrAlaLeuGIuInsSerAlaAlaGIu 426
Db 1360 GAACGTATGATGGTGTGGCTGCACCTTATTAAGAAATATGCTTCTCATGTGTGAAGCAGGA 1419
Qy 427 LysAspAlaAlaLysGIuIleAlaTyrPheLysAspLysLeuHisIleTyrTyrGIu 446
Db 1420 AAGGAAGCTCGAGAAAGTACGGTATTAAGCAAACTTCGCAATATTTATATACAG 1479
Qy 447 AsnSerIleAspAspArgLeuLeuValGIuArgIlePheAlaGIuTyrMetValProHis 466
Db 1480 AACGACATTCAGCAACAACTTGATGAGAAATCTTTGCTGATATCTTGTCCCCAC 1539
Qy 467 AsnLeuGIuThrThrGIuArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
Db 1540 AACCTGGAACAGAAAGAGAGATGAATGCTATATTAATGCTATGCTAGTTGGATCCA 1599
Qy 487 AsnAlaValLysAlaLeuAsnGIuMetTyrPlyCysGIuAsnLeuLeuAlaGHisGIuVal 506
Db 1600 AATCTGTAAAGCTCTCAACGAAATGTGAAGTGAACAATGCTTCGAGGCCATGTA 1659
Qy 507 LysAspLeuLeuAspLeuIleLysGIuProLysThrAspAlaSerValAlaIlePhe 526
Db 1660 CGGCAATATTTGATTTGGTGCACACAGCCATACACAGGCTAATCTTGTCCATCTTT 1719
Qy 527 SerLysValMetValIleThrArgAsnLeuProAspProGIuLysAlaGIuAspPheMet 546
Db 1720 GGAAGAACTGATGACATAGCAAGAAATTTCCCTGACCCCGGAAAGCACAAAGATTTGTG 1779
Qy 547 LysLysPheThrGIuValLeuGIuAspAspGIuLysIleArgLysGIuIleGIuValLeu 566
Db 1780 AAGAAATTTAACAGGTTCTCGCGCATGATGAGAACTTGGTCTCAGTTGGAGTTATTA 1839
Qy 567 ValSerProThrCysSerCysLysGIuAlaGIuIleCysValArgGIuIleThrLysLys 586
Db 1840 ATTAGCCCAACCTGCTTCTTGCAAAACAGCAGATTTTGTGTGAGAGAAATAGCCGGAAA 1899
Qy 587 LeuGIuLysnProLysGIuProThrAsnProPheLeuGIuMetIleLysPheLeuGIu 606
Db 1900 CTTGCAAACTCTTAAGCAACCAAAATCTTTTCAAGATGTCGCAAAATTTCTGTGAA 1959
Qy 607 ArgIleAlaProValHisIleAspThrGIuSerIleSerAlaLeuIleLysGIuValAsn 626
Db 1960 AGAATCGCACCTGTGCATTTGATTCAGAAACCTTAAGTGCACATGAGAAATTCATAT 2019
Qy 627 LysSerIleAspGIuThrAlaAspAspGIuAspGIuGIuValProThrAspGIuAlaIle 646
Db 2020 AAGTCATATAGAGGGGACAGCAGATGATGAGAGAGGGGTAAAGTCCAGATACAGTATC 2079
Qy 647 ArgAlaGIuLysGIuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
Db 2080 CGTTCAGAGCTTGAACTTTAAGCTTCTGCTTTTACAAATCCATCCTGCTGTTCCACTCT 2139
Qy 667 AlaGIuThrPheGIuSerLeuLeuAlaCysLeuLysMetAspAspGIuLysValAlaGIu 686
Db 2140 GCAGAGCATATGATGCTCTTGTATACAGTGGCTTAAGATGAGAGATGCAAGATAGCGAA 2199
Qy 687 AlaAlaLeuGIuIlePheLysAsnThrGIuSerLysIleGIuGIuAspPheProHisIle 706
Db 2200 GCTCTATTCAAATTTTAAAGATACAGGTGCACAAATTAAGAACAGACTCCCCAGATA 2259
Qy 707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysGIuGIuProProAlaGIuAla 726
Db 2260 CGATCGACCTTAATTTCCATTTTAACATCAAAAGCAAGAGGGGTACTGCACACCAAGCA 2319
Qy 727 LysTyrAlaIleHisCysIleHisAlaIlePheSerLysGIuThrGIuPheAlaGIu 746
Db 2320 AAACAGGCTGTGCATCTATACAGCCCATATTTACAAATTAAGAGATCCAGCTTGCACAG 2379

QY	747	LEPHEGLUPROLEHNIHSYSErLEuASPROSeRAsnLEuGLuNIHSLEuLLEPhro	766
Db	2380	ATTTTGGGACCACTAGAGAGTTCTGATCTGATGTGCCAAGAAACCTTATATACCTCA	24339
QY	767	LEuVALThrILEGLIYHISILEALaleuLEuLAPROAspGlnPheALALAProThryLys	786
Db	2440	TTAGTTTCATGGGCGACATTTTCATATGTATACACCAAGATCAAGTTTGGTTCGCCAATGAA	24992
QY	787	SErTPRyVALAThRPhELleVALlySAsPLeuEUKetAsnAspATyLeuProGLyLys	806
Db	2500	TCTGTAGACGCAAAATTTTATGTGTGAAGATCTGTCAAAAGAAATGACAGGTCAACAGGTGA	25598
QY	807	LYSThRThrLYSLeuTPRyVALProAspGLuGLuVALSerProGLuThMEtVALySILE	826
Db	2560	AAGAATGAAAAACCTGTGCTCCAGATTAAGAGTTTCCCTGAAAGTCTACCAAAAGTA	26139
QY	827	GLuNALleLYSMetMetVALArgTPRLeuGLyMEtLYSAsnAsnHISerLYSSEr	846
Db	2620	CAGCCAAATTAACCTTCTGTGAAGGCGCTGTGGATGAAAAACACACACTCTAAATCT	26797
QY	847	GLYhRSErThrLEuATyRLeuLEuThRhrILEuNIHSerAspGLyASLeuThRGLu	866
Db	2680	GCCAAITTCACCCCTTCGGTTATTCACAGCATGTGGTGTAGTGAAGGGGAGCTGTGACAG	27339
QY	867	GLuGLyLYSILESerLYSProAspMetSErATyRLeuATyRLeuALALAGLYSerALILE	886
Db	2740	CAAAAGAGAGATCAGTAAATCTGATATGTCCTGCTGGCATTAAGCTGCGTGAATGCCATA	27992
QY	887	VALlySLeuALAGLInGLuPRocysTyRNIHsGLuLEIethRLeuGLuGlnTyRGLuLeu	906
Db	2800	ATGGAAGCTGTCTCAGGAACCTGTGTACCATGAATATTTACCCCAAGAACCTGTTCAGCTC	28559
QY	907	CysALaleuALALEAsnAspGLICysTyRGLuVALArgGLuVALPheALAGLInLYSLeu	926
Db	2860	TGTGCACCTGTTATTATTAATGATGAGGTATCCAAAGTAAAGCGCATATTTGCTCAGAACTG	29139
QY	927	HISlySGLYSErATyRLeuATyRLeuDPROLeuGLuTyRMEtVALILECysALALECys	946
Db	2920	CATTAAGGCACTGTGGAAGTACTCTGCTCCATGTGGAGTATATGGCGATCTTTCCTTGTGT	29797
QY	947	ALAlYsAsPpROVALySGLuATyRATyRGLuHISALeAGLInCysLEuVALySAsnILE	966
Db	2980	GCCAAAGATCTCTGAGAGAGAGAAAGACACACGACACATGTTTACTGAATAATATTC	30339
QY	967	AsnVALATyRATyRGLuTyRLeuLYSGLuHISALeALeVALSerGLuLYSLeuLEuSErLeu	986
Db	3040	AGTATACGCGACGAATACATTAAACGACAGATCCCTATGGCTACTGAGAAATTTTATCACTG	30939
QY	987	LEuPRocGLuTyRVALyVALProTyRThrILEHISLeuLEuALAHISAsPpROAspTyRVAL	10066
Db	3100	TTTGGCTGAATATGTAGTTCATATCATGTGATTCACCTGCAAGCCATGATCAACGATTTTTCACA	31539
QY	1007	LYSVALGLuAsPpILEGLuGLuLEuLYSAsPVALySGLuTYRLeuTPRPhVALleuGLu	10266
Db	3160	AGATCACAGAATGTGTGATCACTGCTGTATATTCAAAGTGCCTTATGTTTATGCTGTAA	32139
QY	1027	ILEuEUKetALyLSAsnGLuAsnAsnSErHISALePheILEArgLYSMetVALGLuAsn	10466
Db	3220	GTTTAAATATGACAAAGATGAAAAACAAATAGCCATGCTTATGAGAAAGATGCGAGAGAAC	32797
QY	1047	ILEYsGLuThRThrLYSAspALAGLInGLyPROAspAspALySMEtAsnGLuLYSLeuTyR	10666
Db	3280	ATTCAGAGTTTAAACCAAGATGCCCACTCTCCACATGAATCAACAGACAAAGAAACCTGTAT	33339
QY	1067	ThrVALCysAspVALALeMEtAsnILEIleMEtSErLYSSErThRThrTYRLeuGLu	10866
Db	3340	ACAGTATGTGATGTGGCTCTGCTGTATTAATAGTAAAGTGGCTTGTGCAATGACAGAT	33939
QY	1087	SErPRocLYSAsPpROVALleuPRoALeArgPhePheThRGLInPRoAspLYSAsnPheser	11066
Db	3400	TGACCAAGAGCCCAAGTCTCTCCATGAATATTTTATACACAACCTGAAAGAGACTGTGTGT	34539

[illegible]

PH Key Location/Qualifiers
 FT CDS 140..2704
 FT /*tag= a
 FT /product= Neuron associated protein
 XX
 PN MO200034477-A2.
 PD 15-JUN-2000.
 XX
 PF 10-DEC-1999; 99MO-US30408.
 XX
 PR 11-DEC-1998; 98US-0210083.
 PR 11-DEC-1998; 98US-9123456.
 PR 09-FEB-1999; 99US-0119365.
 PR 16-MAR-1999; 99US-0124687.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
 PI Lu DM, Azimzal Y;
 XX
 DR WPI: 2000-423423/36.
 DR P-PSDB; AAB01382.
 XX
 PT New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders
 XX
 PS Claim 9; Page 130-131; 145pp; English.
 XX
 CC Human neuron-associated proteins (NEUP) can be used for for
 CC treating or preventing a disorder associated with decreased
 CC expression or activity of NEUP. Antagonists of NEUP are useful for
 CC treating or preventing disorder associated with increased expression
 CC or activity of NEUP. NEUP or their fragments or derivatives are
 CC useful for treating neurological disorder such as epilepsy, ischemic
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
 CC disease, Pick's disease, Huntington's disease, dementia and
 CC Parkinson's disease. NEUPs are also useful for treating other
 CC demyelinating diseases, bacterial and viral meningitis, prion
 CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
 CC other neuromuscular disorders, peripheral nervous system disorders,
 CC inherited, metabolic, endocrine, and toxic myopathies, mental
 CC disorders including mood, anxiety and schizophrenic disorders, a cell
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 CC hemoglobinuria, cancers of the adrenal gland, bladder, bone,
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 CC disease, adult respiratory distress syndrome, allergies, ankylosing
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
 CC complications of cancer, hemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This sequence was given the Incyte ID no.
 CC 2888437C81.
 XX
 SQ Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;
 Alignment Scores:
 Pred. No.: 1 66e-194 Length: 3957
 Score: 2932.50 Matches: 595
 Percent Similarity: 74.51% Conservative: 133
 Best Local Similarity: 60.90% Mismatches: 185
 Query Match: 40.77% Indels: 65
 DB: 21 Gaps: 16
 US-09-512-581b-2 (1-1391) x AAA47423 (1-3957)

QY 428 ASpAlAlAlAlScInIleAtATpIleLysAspLysLeuLeuHISileTyrTrgInAsn 447
 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 2 GAAGCTGACGAGAAAGACACCTGGATAAAGCAAACTCTGATATTATATATATACAGAAC 61
 QY 448 SerIleAspAspArgLeuLeuValGluArgIlePheAlaGluPheMetValProHisAsn 467
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 62 AGCATTTGACGACAAACCTGTTGGTAGAGAAATCTTGTCTCATATCTGCCCCCAAC 121
 QY 468 LeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 122 CTGGAACACAGACAGAAATGAATGATCTTATATTACTTATATCTGATTGGATTCGAAT 181
 QY 488 AlaValLysAlaLeuAsnGluMetTyrPlyCysGlnAsnLeuLeuArgHisGlnValLys 507
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 182 GCTGTAAACAGCTCTCAACGAAATGTGAAGTGTCAACATCTGCTGATCCATGTACCG 241
 QY 508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 242 GAACCTATGTGATTGGACAAAGACGCTACATTCAGAGCTAACCTTCTGCTGCAATTTGGA 301
 QY 528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 302 AAACGTATGACCATGACAAAGATTTTGCCTGACCCCGGAAACACAGATTTTGGAGAC 361
 QY 548 LysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLysVal 567
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 362 AAATTTACCAAGCTTCTCGGCGATGTAGTGAACAACTCGCTCATGTTGAGTTATTAAT 421
 QY 568 SerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu 587
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 422 AGCCCAACCTGTTCTTGCACAAACAGAGATTTTGTGTGAGAAATAGCCCGGAACCTT 481
 QY 588 GlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPheLeuLeuGluArg 607
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 482 GCAATCTTACGACCAACCAAAATCTTTCTAGAGATGTGCAATTTCTGTGGAAGA 541
 QY 608 IleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 542 ATCGACACTGTGACATGTGATTCAGAACCATAGTGCACCTGATGAATGAATGAATAG 601
 QY 628 SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArg 647
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 602 TCATATAGAGGGGACAGACAGATGATGAAGAGAGGGTGTAGTCCAGATACAGTATCCGT 661
 QY 648 AlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAla 667
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 662 TCAGACCTTAATCTTAAAGTGTCTGCTTTTAAACATCTTCACCTGCTCCACTGCA 721
 QY 668 GluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAla 687
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 722 GAGACATATGAGCTCTGTTCAGTGCCTTAAGATGAGATGACAAAGGTAGCAGAAAGCT 781
 QY 688 AlaLeuGlnIlePheLysAsnThrGlySerLysIleGluLysPheProHisIleArg 707
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 782 GCTATTTAAATTTTAAATAATACAGGTCAACAAATTAACAGACCTTCCCAAGATACGA 841
 QY 708 SerAlaLeuLeuProValLeuHisHisLysSerLysGlyProProArgGlnAlaLys 727
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 842 TCGACCTTAATTCATTTTACATCAAAAAGCAAAAGGGGTACTCCACACCAAGCAAAA 901
 QY 728 TyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIle 747
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 902 CAGGCTGTGCATGTATACAGCCCATATTCACAAATTAAGAAAGCTTCAGTCACAGATT 961
 QY 748 PheGluProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrProLeu 767
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 962 TTTGAGCACTCTCAGTAGCTGGAAGTGTGATGTGCCAACAAACTTATTAACCTCAATTA 1021
 QY 768 ValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaIleProTyrPlySer 787
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1022 GTTTCATTTGGCCACATTTTCTATGTTAGCCACAGATCATGTTTGTCTCCCAATGAATCT 1081
 QY 788 TrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807

XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PMD, Myers EW,
 XX WPI; 2001-656860/75.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 17788; 21bp + Sequence Listing: English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB101840-AB116175).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3920 BP; 1053 A; 1007 C; 947 G; 913 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,42e-128 Length: 3920
 Score: 1986.50 Matches: 448
 Percent Similarity: 57.77% Conservative: 258
 Best Local Similarity: 36.66% Mismatches: 441
 Query Match: 27.62% Indels: 76
 DB: 23 Gaps: 22
 US-09-512-581b-2 (1-1391) x AB122105 (1-3920)
 QY 13 IIEThTyrProProGluValLysGluLLeSerAspLysLLeSerLysGluGluMetVal 32
 DB 273 ATAGTACACCCACCGGATGCGCGCGCTGTGAGAGATTGGGACAGACGAGCTGATA 332
 QY 33 ATATGLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluGlu 52
 DB 333 AGCGGCTCCAGACCCCTCGCCAAATGTTCTGACAGACCAATGAGCAGCAATCTGTAC 392
 QY 53 GLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 72
 DB 393 CAGCAAA---TACATACCATTTGGCCCTCCATCTGTCGACGACATCTTTATGCAACATCCA 449
 QY 73 GLeuLysAspValArgLeuLeuValAlaCysCysLeuAlaAspLLePheArgLLeTyrAla 92
 DB 450 TCCCGGATGTCAGATCTTAAATGCTTAAATGCTTAAATGCTTAAATGCTTAAATGCTTAA 509
 QY 93 ProGluAlaProThrSerProAspLysLeuLysAspLLePheMetPheLLeThArg 112
 DB 510 CCGAGGCTCCCTACAAAGCAAGCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 569
 QY 113 GLeuLysGLeuLysAspThrLysSerProGluPheAsnArgTyrPheTyrLeuLeu 132
 DB 570 CAACTCAGCGCTGAAGATCGCGCGATCTCTTTCAGAGCTTACTCTATCTCTG 629
 QY 133 GLeuLysGLeuLysAspThrLysSerProGluPheAsnArgTyrPheTyrLeuLeu 152
 DB 630 GAGAACCTGCTTCTTCAAGCTTAAATGCTTAAATGCTTAAATGCTTAAATGCTTAAATG 689
 QY 153 ILePheThrGLeuLysArgThrLysPheSerValLLeAsnAsnGlnLys 172
 DB 690 ATCTTCAAGACCTGTTTACTACATCTCAAGATTCATCAAGATTCATCAAGATTCATCAAG 749
 QY 173 ValHisMetHisMetValAspLeuMetSerSerLLeLLeCysGluGluAspThrValSer 192
 DB 750 GTGACATACCTTTTCTAGATGTGCTAAGCGCTTATCATCAGACGCGCAATCTGTGCG 809

QY 193 GLeuLysLeuLysAspThrValLLeuValAsnLeuValProAlaHisLysAsnLeuAsnLys 212
 DB 810 GTGAAATGTTGTGATCATCTTATGATTAACATGATGAGCGCTTAACAAATCTTACAAACAG 869
 QY 213 GLeuLysAspLeuLysAlaLeuLysArgThrLLeGlnAlaLeuGluProTyr 232
 DB 870 TTTGCTCGCAATTAACAGAACACTTCTACTTAAAGGGCGCATGCTCTCGAGTCCACC 929
 QY 233 ILeThrThrPheAsnGlnValLeuMetLeuGluLys-----ThrSerLLeSerAsp 250
 DB 930 ATCAAAATGTTCTTAAACGCTGCTGTGATGATGATGATGATGATGATGATGATGATGATG 986
 QY 251 LeuSerGluHisValPheAspLeuLLeuGluLeuTyrAsnLLeAspSerHisLeuLeu 270
 DB 987 ATTAACAACAAATCTTATGATTAACATGATGATGATGATGATGATGATGATGATGATG 1046
 QY 271 LeuSerValLeuProGluGluGluPheLysLeuLysSerLLeAsnAsnGlnLysGluGlu 290
 DB 1047 TGTTCAGTCTACCTCAGCTGAGAACAACTGCTGCCAGACGATGCTGAGAGGCTA 1106
 QY 291 GLeuValValLysLeuLeuAlaLysMetPheGluValAlaLysAspSerGluLeuAlaSerGln 310
 DB 1107 AAGGCAACCACTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1166
 QY 311 AsnLysProLeuThrGlnCysTyrLeuGlyArgPheAsnAspLLeHisValProLLeArg 330
 DB 1167 TACCCCAACCTGTGAATAATCTTCTTGTGCGCTTGTGCGACATCACCGAACAGATTGCG 1226
 QY 331 LeuGluCysValLysPheAlaSerHisCysLeuMetAspProAspLeuAlaLysAsp 350
 DB 1227 ATCAAGGTGTGTATGATCATCATGATCATCTCTGCTCATATCAACCACTTCCAGCAGAT 1286
 QY 351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaLeuAlaGlnHisAspVal 370
 DB 1287 ATCAACCGAATAATTTGGCGCTTCGAATATCATCATGATCATGACAGATGCTGCGCAGAGTG 1346
 QY 371 ILeValSerLLeValThrAlaAlaLysLysAspLLeuLeuValAsnAsp-----His 388
 DB 1347 GTAATGCTATTTTGTGAATCTTGAACGCGGATGATGATGATGATGATGATGATGATGAT 1406
 QY 389 LeuLeuAsnPheValArgGluArgThrLeuAspLysArgThrArgValArgLysGluAla 408
 DB 1407 CTACTGAATAATTTGCGCGGAGCGGAGCGGATGATGATGATGATGATGATGATGATGATG 1466
 QY 409 MetLeuGLeuLysLeuAlaGlnLLeTyrLysLysTyrAlaLeuLeuSerAlaAlaGluLysAsp 428
 DB 1467 ATGAATGTGTGGCTTACATCTTAACAGCGGCAATTTGCAACCAAGATCAAGCACT 1526
 QY 429 AlaAlaLys---GlnLLeAlaThrLLeLysAspLysLeuLeuHisLLeTyrGluAsn 447
 DB 1527 GGCCTTAAGGTCAGGCTGATGATTAACAAAGATCTTACATGATGATGATGATGATGATG 1586
 QY 448 SerLLeAspAspArgLeuLeuValGluArgLLePheLLeGlnLysMetValProHisAsn 467
 DB 1587 GGCCTGAGGAGCGGCTGCTTGTGAGAGGCTTATATCACTGCTGCTGCTGCTTAA 1646
 QY 468 LeuGluThrThrGluArgMetLysCysLLeTyrLysTyrAlaThrHisLeuAspLeuAsn 487
 DB 1647 CTAGCTCCGGAAGAGCGCATGAAGATCATATCATATTTGCTAGCGATTCGATGCCAAT 1706
 QY 488 AlaValLysAlaLeuAsnGluMetThrLysCysGlnHisLeuLeuArgHisGlnValLys 507
 DB 1707 GCCACCAAGGCTTCTGCTGAGCGGAGAACAAATGAAGACAGCAATAGCAATAGCAATAG 1766
 QY 508 AspLeuLeuAspLeuLLeuLysGlnProLys---ThrAspLLeSerValLysAlaLLePhe 526
 DB 1767 GATTGGATCAAGCTACACCACTTAAGAGATTCACGCGCGCTACTTACTACGCTCAGC 1826
 QY 527 SerLysValMetValLLeThrArgAsnLeuProAspProGluLysLLeGlnAspPheMet 546
 DB 1827 GCCAAGAGGCGCAACATTTGCCAAGATGCTTCCAGATTCCTTAAGAGGAGAGTACCTA 1886
 QY 547 LysLysPheThrGlnValLeuGluAspAspGluLysLLeArgLysGlnLeuGluValLeu 566

Db	1887	ACCAGCTTTAGACAAACAAACGATGCTGAGATGCTGATCAAGATGCAATTAAACATTGCTC	1946
QY	567	ValSerProThrCysSerCysLysGlnAlaGluCysValArgGluIleThrLysLys	586
Db	1947	CTTAAACGCGACGTAAAGCTGTGCGGAGAGTGTCTGCACAGAGGAGAGTTCTCTCGTGAAGAAG	2006
QY	587	LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLeuPheLeuGlu	606
Db	2007	CTTGGCGGCCACAGTCCAA---TCGATGTGATTAAACACAGATTAAAGATGCTGATTGAG	2063
QY	607	ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn	626
Db	2064	CGCGGGCATCATGATGCTGAGACAGAGAGCTATTGTGCGTCAATTTAGCCCTAAATTGAA	2123
QY	627	LysSerIleAspGlyThrAlaAspArgLysGluValProThrAlaProThrAspGlnAlaIle	646
Db	2124	CAATGTCATCGAAGAAAGGTCATTTGTCATGAAGATGCGAATTTGGCTCAGAGAAAGGGGC	2183
QY	647	ArgAlaGlyLeuGluLeuLysValLeuSerPheThrHisProIleSerPheHisSer	666
Db	2184	GAGCGCGGACATCAAGTTGTGGCATGCTTTCATACAGCTTTTGGCGCTCATTTCTGCACC	2243
QY	667	AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspArgGluLysValAlaLeu	686
Db	2244	GATACTCTGCGCTCATTTAAATTTTCCCTGCTGATTAAGATGCGAAGATTATGTTGACACA	2303
QY	687	AlaIleLeuGlnIlePheLysAsnThrGly-----SerLysIleGluAlaPhe	703
Db	2304	CTGGTGTCAAAAGACATACATCTGTGGGGCGGCTATACGCCATCTATTGACAGCCCTCA	2363
QY	704	ProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLysGlyProPro	723
Db	2364	CCGGCTATTCTCGACAGAGCTGCGTCCGGGTGCAAGACATCGCATATTGGAACCC	2423
QY	724	ArgGlnAlaLysTyrrAlaIleHisCysLienHis-----	734
Db	2424	AAGCAAGCAAGACACGCGGTCCGATGTATTTTCTGAACAGCCAGCTGCGGCTCCACT	2483
QY	735	-----AlaIlePheSerSerLysGluThrGln-----	743
Db	2484	GATGAGACAAACATGAGCTGGAAGTGCATGCCACATCAACAAACAGTCACTCTCTT	2543
QY	744	PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeu	763
Db	2544	TTCAACGATCATCGACGCTTGCGCTTAACATGACACACA---AACTGTGACATCAG	2600
QY	764	IleThrProLeuValThrIleGlnHisIleAlaLeuLeuAlaProAspGlnPheAlaAla	783
Db	2601	CGCACAAGATGTGAGACTGGGTGCACATGCTCAATATGCCACAGGCTTCTCTAAG	2660
QY	784	ProThrLysSerThrPvalAlaThrPheIleValLysAspLeuMetAsnAspArgLeu	803
Db	2661	CCCATTTAAGATATGATGATGCGCGAAGATGTCCAAGAGCGCTTATGCCAAGAA---GTT	2717
QY	804	ProGluLysLysThrThrLysLeu-----ThrValProAspGluGlnVal	818
Db	2718	CTTGGCGGACAGACTAGAACTGCCAGAGACAGTATGTGTGTGCCCAAGAGAACTA	2777
QY	819	SerProGluThrMetValLysIleGlnAlaIleLysMetValArgThrPheLeuGly	838
Db	2778	CCGCGCGACACTATGTAAGCTGGAATGCGCTCAAGAGGTATGCGCAGGTGCTTTGGGA	2837
QY	839	MetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeu	858
Db	2838	TTTGGTACCGAT---GAGCAGCGCTCCCGAAGAACTTCCGAATGCTGGCGGCTTTGTT	2894
QY	859	HisSerAspGlyAspLeuThrGlnGlnLysLysLeuLysProAspMetSerArgLeu	878
Db	2895	AACCAACAGAGCGATTTCTGGCGCAAGAACGCTTTTCCGCTGCCGAGAAATCTTGGCTG	2954
QY	879	ArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrrHisGluLeu	898

Db	2955	CGCCCTGGGGGACAGCTGGCGCATGTCTCAAAGTGTCCGAACAAAGAGCGGTAGGTATCAG	3014
Qy	889	ILethrleuglnGlnTyrGlnLeucCysalaIeualaIleasnAspGluCysTyrGlnVal	918
Db	3015	TACACGCGTGAAGCACTATTTCAGCTTTCACAGCTGATGGCTGACCGGTCCGAGAACTT	3074
Qy	919	ArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArg-----	934
Db	3075	CGGGAAATCTTTGGCTGGCAGAGCTGCACAAAGGATTAAACAGAAAGTTTCCACGAACTGT	3134
Qy	935	LeuProLeuGlnTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlnArg	954
Db	3135	TTGGCGGTGAGACTTCATGGCGCTTGTATGGTGGCGGTGTGTAGACACTGAGAGGAAATTGG	3194
Qy	955	ArgAlaHisAlaIleArgGlnCysLeuValLysAsnIleAsnValArgArgGlnTyrLeuLys	974
Db	3195	CAGAAGCTGTGGCGCACTATGACAGAAAGGATGTAAACAAAGCGGGAAATATCTCAAG	3254
Qy	975	GlnHisAlaAlaValSer-----GlnLysLeuLeuSerLeuLeuPro	988
Db	3255	ACTGTGGGTATGACATCTTCGCCGACAGCTCAACGGAGTGCACATATCTTACATACTACTCT	3314
Qy	989	GlnTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysVal	1008
Db	3315	GACTCAATGCTGGCGTTTGGCTATTCGCCGTGGTGGCCAGATCCAGCTTCACGAAATCAC	3374
Qy	1009	GlnAspIleGlnGlnLeuLysAspValLysGlnCysLeuTppPheValLeuGlnIleLeu	1028
Db	3375	GAGGAGCTACGTACACTCTGGCAGATGGAGAGATCCCTGGCTTATCTGTGAGCCGCTG	3434
Qy	1039	MetAlaLysAsnGlnAsnAsnSerHisAlaPheIleArgLysMetValGlnAsnIleLys	1048
Db	3435	ATGGCCAAACGAGAAACGTTTGTCCATGTACTTCTCAAGACAGCTCTGACACTGTAAAG	3494
Qy	1049	GlnThrLysAspAlaGlnGlyProAspAspAlaLysMetLeuGlnLysLeuTyrThrVal	1068
Db	3495	CATCCGGAGTTTCAGTCTGGGG---TCGGACAAAGCGCGACACATATAAATGTGGCGCTC	3554
Qy	1069	CysAspValAlaMetAsnIleIleMetSerLys-----SerThr	1081
Db	3552	TGCCATCTTGGCATCTACATATATCATGACCTCAAGTCCACTTCATTTGATGGCAACAGAGC	3611
Qy	1082	ThrTyrSerLeuGlnSerProLysAspProValLeuProAlaArgPhePheThrGlnPro	1101
Db	3612	ACCTTTTCCATGCGCGCTG-----GCTTGTCCAGAAATGATTATTAAGAGCTT	3655
Qy	1102	Asp---LysAsnPheSerAsnThrLysAsnTyrLeuProGlnMetLysSerPhePhe	1120
Db	3660	GCCGTGGGAATTTCCAAAACATGACGTATATATACCGCTGAGAGCTG-----TAT	3710
Qy	1121	ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer	1140
Db	3711	ACGCGGGAGAGCCAAATTCACAGGAGCAAAAGCTGCCGACAGC-AATGACAAAGCTCCGAGC	3765
Qy	1141	AlaGlyLysGlnSerGlnThr-----LysSerSerArg	1151
Db	3770	AGCAATGGCTCCAAAGAGAACCGCGCGAAGCACTGATCATGTAGTAAATCCCGACAGA	3829
Qy	1152	MetGlnThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLys	1171
Db	3830	GAAACAATCTGTTGCGACACATACAGAGCGGCTGACACTACGAGAGCCCATGGCCAAACGAAC	3889
Qy	1172	GlyArg 1173	
Db	3890	GCGCGC 3895	
RESULT 10			
ABL22104/c			
ID ABL22104 standard; DNA: 6950 BP.			
AC ABL22104;			
XX			
DT 26-MAR-2002 (first entry)			

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 17785.
 XX XX Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PMD, Myers EW;
 XX DR WPI; 2001-656860/75.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Claim 1; SEQ ID NO 17785; 21bp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 6990 BP; 1907 A; 1513 C; 1586 G; 1984 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,48e-110 Length: 6990
 Score: 1729.50 Matches: 463
 Percent Similarity: 49.43% Conservative: 270
 Best Local Similarity: 31.22% Mismatches: 436
 Query Match: 24,04% Indels: 318
 DB: 23 gaps: 37
 US-09-512-581b-2 (1-1391) x AB122104 (1-6990)
 QY 13 IletThyrProProGlyValLysGluIleSerAspLysIleSerLysGluIleMetVal 32
 DB 5398 ATAGTGTACCCACCCGATGCCGCGCTGTGAGGATTTGGACACAGCAGCTGATA 5339
 QY 33 ArgArgLeuLysMet-ValVal----- 39
 DB 5338 AGCGCGCTCAAGGTGAGTCTCGCCCTGTTCCTTAGCAAAACTTCTACTAGATGA 5279
 QY 40 -----LysThrPheMetAspMetAspGlnAspSerGluGlu 52
 DB 5278 TTCTGTTTACAGACCTCGCCCATGTTCTGCAGACCATGACAGCAGCAATCTCTA 5219
 QY 52 ulysGluLeuThyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
 DB 5218 CCAAGCA---TACATACCATGTGGCCCTCCATCTGCTGGACGACTTCTTATGCAACATCC 5162
 QY 72 ogLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAl 92
 DB 5161 ATCCGATGATGCTCAGACTTATATCCGCTCTCTCGTGGCGAGATGCTCTCGGGGTTATGC 5102

QY 92 aProGluAlaProTyrThrSerProAspLysAspIlePheMetPheIleThrAr 112
 DB 5101 GCCAGAGGCTCCCTACAAAGACAGACATCAAACTATATTATTTATTTATCA 5042
 QY 112 gGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeu 132
 DB 5041 GCACCTGCAGGCTGTGAAGATCCGCGCATCTCTTCAACCGTACTCTATCTGCT 4982
 QY 132 uGluAsnIleAlaTyrPheLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn 152
 DB 4981 GGAGAACCTTGCTTTTCTCAAGTCTTTATATGTCTTTTGAACGTGAGACTGCCAAG 4922
 QY 152 uIlePheThrGlnLeuTyrArgThrLeuPheSerValIle----- 165
 DB 4921 AATCTTCAAGACCTGTTAGTACACATCTTCAAGATTTGCAAGTACGAGCTTCTCT 4862
 QY 166 -----AsnAsnGlyHisAsnG 171
 DB 4861 TTATCTGCGAAGCTTTTATATATCAGATGTATATATTTTCTAGTACAGCAGAGTGT 4802
 QY 171 nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVa 191
 DB 4801 CAAGGTACTACTTTTCTTCTAGATGTGCTAAGCCGTGTATCAGACAGACCGCAATCT 4742
 QY 191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
 DB 4741 GTGCGTGAATGTGGATCTCATTTGATTAACATGTAGAGCCGCTCAAACTTAACA 4682
 QY 211 nLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGlu-P 231
 DB 4681 CAAGTTTGCTGCCAATTACAGAACACTTCTCACTAAACGAGCGGATGCTCTCAGATC 4622
 QY 231 fo-----TyrIleThrThr----- 235
 DB 4621 CACCATCAAAATGGTATATTGCTGTGATCTTTGGTGCACACAGTTATTCACCTAATCC 4562
 QY 236 -----PhePheAsnGlnValLeuMetLeuGlyLys-----ThrSerIleSera 250
 DB 4561 CTGATCTACAGCTCTTATTAACCGTCTGCTGATGATGATGATGATGATGATGATGAT 4503
 QY 250 spleuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu 270
 DB 4502 --ATTACAAACAAATTTATGATATATTTTACAGAGCTTAACCCATCAATGCAAGCTTGT 4445
 QY 270 euleuSerValLeuProGlnLeuGluIlePheLysSerAsnAspAspGluArgL 290
 DB 4444 TGTGTTAGTGCTACCTGACCTGAGAACAACTGCTGCACAGAGATGCTGAGAGGC 4385
 QY 290 eu----- 290
 DB 4384 TAAGTAGTATTATATCAGCATGTACCTCTATTGTCATATTATTAACATATTTCGTTA 4325
 QY 291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAla 309
 DB 4324 CTACAGAGGCAACACTCTCTGTGCTGTGATGTTCGCCAAAGAGATGCCAATCTAGCCA 4265
 QY 309 eArgAsnLysProLeuTyrGlnCysTyrLeuArgPheAsnAspIleHisValPro 329
 DB 4264 AAAAGTACCCCAACCTGTGAAATGCTTTTGGCGCGCTCTCGACATCAGCAACGACAG 4205
 QY 329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 349
 DB 4204 TTGCAATCAAGTGTGTCATCTCAATGCAATGCACTTCTGCTCAATCAACCCAGTCTCAGC 4145
 QY 349 ysAspLeuThrGluTyrLeuValArgSerHisAspProGluGluAlaIleArgHisA 369
 DB 4144 ACGATATACCCGGAATTTGCGGCTTGAATATACGATCTTTCAGAGATGTGTCGCGACG 4085
 QY 369 sPValIleValSerIleValThrAlaLysLysAspIleLeuLeuValAsnAspP--- 387
 DB 4084 AGGTGTGAATAGCTATTGTGGAACATGCAAGGCGACCTCACCTGCTGCGAAGCGC 4025
 QY 388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTyrArgValArgLysG 407

Db 4024 CCGATCTACTAGAAATTTGGCCGAGCGAGCCTAGATAAGAAATTCACAGG 3965
Oy 407 LuAlaMetMetGlyLeuAlaGlnIleTyrlsLysTyrlAlaLeuInserAlaIaGlyL 427
Db 3964 ATGCATCAATAGGTGGCTTACATCTACAAACCGCAATTTGCGCAACCAAGATCTAA 3905
Oy 427 yAspAlaAlaLys---GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrlTyrg 446
Db 3904 GCACGGCCTTAAGGTACAGGTTGACTGATTAAACAAGATCTACATGATCTACA 3845
Oy 446 lAsnSerIleAspAlaTrpLeuLeuValGluArgIlePheAlaGlnTyrlMetValProH 466
Db 3844 AAGTGGCTTGAGAGCCGCTGCTTGGAACGCTTACTTATCACTGCGCTGTTCCCT 3785
Oy 466 lAsnLeuGlnThrThrGluArgMetLysCysLeuTyrlTyrlLeuTyrlAlaThrLeuAspL 486
Db 3784 ATAACTTACGTCGGGAAAGCGCATGAAGCTTATTCATTTGCTAGCGCATTCGATG 3725
Oy 486 euAsnAlaValLysAlaLeuAsnGlnMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
Db 3724 CCATGCCACCAAGGCTTTCGTCGAGCTGCAGAGAACCAATGAAGACGCAATACG 3665
Oy 506 aLysAspLeuLeuAspLeuIleLysGln-----ProLysThrAspAlaSerV 522
Db 3664 TCAGTGTATGTGATCAAGCTACACACCTAAGAGATTCACGCGCGCTACTTACTACG 3605
Oy 522 aLysAla----- 524
Db 3604 TCACGCGCCAGAGCCCAACATTCGCCAAGTGGTTCTATTAATAGTTTCTTTGATAGT 3545
Oy 525 --IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
Db 3544 TTGTAAACAGAGATGTGTTTTTTTCTTCAGACCTCCAGATCCCTTAAGACGGCA 3485
Oy 544 AspPheMetLysLysPheThrGlnValLeuGlnAspAspGlyLysIleArgLysGlnLeu 563
Db 3484 GAGTACCTTACCCAGTTTACGACAAACCTGCGAAAGATGCTCAGCTCCTCAAGATCAT 3425
Oy 564 GluValLeuValSerProThrCysSerCysLysGlnAlaGlnGlyCysValArgLysIle 583
Db 3424 AACATGTCTTAAACGCGAGCATGAGCTGCGGAGGTGCTGCACACATGGAGATTCCT 3365
Oy 584 ThrLysLysLeuGlnLysProLysGlnProThrAsnProPheLeuGlnMetIleLysPhe 603
Db 3364 CTGAAAGAGCTTGGCGCCAGCTCCAA---TCGAATCTGTATTACAAACAGATTAAAGT 3308
Oy 604 LeuLeuGlnArgIleAlaProValHisIleAspThrGlnSerIleSerAlaLeuIleLys 623
Db 3307 CTGATTGAGCGGTGCGCATGATGTGGCAAGAGTCTATTCGGCTATTATAG 3248
Oy 624 -----Gln 624
Db 3247 TAAATGCTTTAAGTCCCTTTTCCAGACTTAATCCGTAATCTTTTGTCCAGCCTA 3188
Oy 625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspL 644
Db 3187 ATTAAACATGATCGAAGAAAGGTCATGTGTGAAGAGATGGAAATTTGCGTTCAGGA 3128
Oy 645 AlaIleArgAlaGlyLeuGlnLeuLeu----- 653
Db 3127 GCGGGCGAGCGCGAGCTCAATGTTGTGGCCCTGAGTATAGATTAATTAAATATGATA 3068
Oy 654 -----LysValLeuSerPheThrHisProLysSer 663
Db 3067 CAGAAATCCATTTGATGCTTATATCTTATCTGACAGATG-CTTCAATAGCTTTTGGCTCAT 3009
Oy 664 PheHisSerAlaGluThrPheGlnSerLeuLeuLacLysLeuLysMetAspAspGluLys 683
Db 3008 TTCTTCACCGATCTGCTGCTGCATTAATTTCCCTGCTCAGTTACAGCAGGATTAAT 2949
Oy 684 ValAlaGluAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700
Db 684 ValAlaGluAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700

Db 2948 GTTGACCACTAGTGTCTAAAGACACTAACTATCTAGGGCGGCTATCAGCCACTATGAC 2889
Oy 701 GluAspPheProHisIleArgSerAlaLeuLeuProValIleuHisIleLysSerLysLys 720
Db 2888 GACCTTACACCGGCTATTCTGACAGAGTGGCTGCGGTGCGAAGACTTGGCACTTAAT 2829
Oy 721 GlyProPheArgGlnAlaLysTyrlAlaIleHisCysIleHis----- 734
Db 2828 GGAACCCCAAGCAAGCAAGACCGCGGTGCATATTATTTGCTAAAGCAGTCGTCG 2769
Oy 735 -----AlaIlePheSerLysGluThrGln----- 743
Db 2768 GCTTCACATGATGAGCAACAAGTGGACTGGAATGCGTCACCAACTAACAACAGATG 2709
Oy 744 -----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2708 CATCTTAATTTTCAACAGATCATGACAGAGCTTCGCTAAACTGACACA---AACGT 2652
Oy 761 GluHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeuAlaProAspGln 780
Db 2651 GAGCATCAGCCGACAAAGATTTGTGACTTGGGTGCATTTGCTACAAATATGCCACAGGC 2592
Oy 781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Db 2591 TTCTTAACGCCCATTAAGAAATATGATTCGCGACAAATTTGTCAAAGACTGCTTATCCA 2532
Oy 801 AspArgLeuProGlyLysLysThrThrLysLeu-----TrpValProAsp 815
Db 2531 GAA---GTTCCCTGCGCGAGACTTACAGACTGCGCAGAGACAGTGAATGTTGTCCTCA 2475
Oy 816 GluGluValSerProGlyThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
Db 2474 GAGAACTACCCGCGGACACTCTATGTAAGCTGATGCGCTCAAGGCTATGCGCCAGTGG 2415
Oy 836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
Db 2414 CTTTGGGATTTGCGTACCCGAT---GAGCAGCGTGGCCGAGAAACATTCGAAATGCTGGCG 2358
Oy 856 ThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMet 875
Db 2357 GCGTTTGTAAACAAAGAGGATTTGCTCGGCCAAGAACGCTTTGGCGGTGCGGAGAA 2298
Oy 876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrl 895
Db 2297 TCTTGGCGCGGCTGCGGCGACGCTGCGCCATGCTCAAGTGTGCGAACAAGCGCTA 2238
Oy 896 HisGluIleIleThrLeuGlnGlnTyrl---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
Db 2237 GGTGATCAGTACAGCGCTGACAGATTTTGCAGCTTTCGCCAGCTGATGTGATGAGAC 2178
Oy 914 ----- 914
Db 2177 ATTCATCTTGTCTTGAACCTTACTGTCATATGTCGCTGCTTCCGTAAGCTGATC 2118
Oy 915 ---CysTyrlGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
Db 2117 CGGTGCCA-GAAGTTGGGAAATCTTGTCTGCAAGCTGCACAAAGGATTTAAGCGAAGT 2059
Oy 934 Arg-----LeuProLeuGlnLysMetAlaIleCysAlaLeuCysAla----- 947
Db 2058 TTGCCCAAGAACTGTTCCGCTGACTTCAATGGCTTGTATGTCTGCGTGGTGTAGAG 1999
Oy 948 -----LysAspProValLysGluArgAlaArgAlaHisAlaArg-GlnCysLeuValLys 964
Db 1998 ACTGAGAGAGAGTAGTGTTCAGAAACCATTCGATCCGTCGAGAAAGTATGATTA 1939
Oy 964 s-----AsnIleAsnValArg 969
Db 1938 TATTTTCAACAGATTGCAAGACTTGTGCTGCTACTATGACAAACGGATTTAAACAAAG 1879
Oy 969 gArgGluTyrlLeuLys----- 974
Db 1878 GCGGGAATATCTCAAGACTGTGCGTATGACATGTAGTGTTTACTTTTGAATCATATT 1819


```

Oy 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrValProAspGlu 816
Db 106 -----ACAGGTGAAAAAGATGGAAGAACTGTGGCTCCAGATGA 144
Oy 817 GluValSerProGlyThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeu 836
Db 145 GAGGTTCCTCCGAGTACTAGCAAGAGTACAGCAATTAACTCTGTGTAAGTGGCTG 204
Oy 837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrTr 856
Db 205 TTGGGTATGAAAAAACACCGACGTCTAAATCCCAATTCAGCCCTTCGGTTATTATTCAGCG 264
Oy 857 IleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMetSer 876
Db 265 ATGTTGGTTAGTGAAGAGGTGACCTGACAGCAAGAAAGAGATCAGTAAATCTATATGTCT 324
Oy 877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTrpHis 896
Db 325 CGCTTGCATTAAGTGTGTGTGTCATTAATGAAGCTTGCTCAGAGACCTTGTTACCAT 384
Oy 897 GluIleIleThrLeuGlnGlnThrLysLeuCysAlaLeuAlaIleAsnAspGluCysTrp 916
Db 385 GAAATTTATTACCCCGCAACAGTTTACGCTGTGTGCACTGTTATTATGATGATGAGTTAC 444
Oy 917 GluValArgGlnValPheAlaGlnLysLeuHisLysGlyLysSerArgLeuArgLeuPro 936
Db 445 CAAGTAAGCAGCATTTTGTCTCAGAAAGCTGCATAAGGCATTTGTGAAGTACTGCTCCCA 504
Oy 937 LeuGlyTrpMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgAla 956
Db 505 TTGGCATATATGGCATCTTTGCTGTGTCGCAAAAGATCCTGTGAAGAGAGAGAGA 564
Oy 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgGlnTrpLeuLysGlnHis 976
Db 565 CACGCACGACAGATGTTTACTGAAAAATATCAGTATACCCAGGGAATTAATTAGCAGAT 624
Oy 977 AlaAlaValSerGlyLysLeuLeuSerLeuLeuProGlyTrpValValProThrTrpIle 996
Db 625 CCTATGGCTACTGAGAAATTTATTTACTGTTGGCTGATATATGTGTTCCATATCATGATT 684
Oy 997 HisLeuLeuAlaHisAspProAspTrpValLysValGlnAspIleGlnLeuLysAsp 1016
Db 685 CACGCTAGCCCATGATCCAGATTTTACAGATGACAAAGATGTTGATCAGCTTCGTGAT 744
Oy 1017 ValLysGlyCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnSer 1036
Db 745 ATCAAAAGAGTCCATATGTTTCATGCTTGAAGTTTAAACAAAGAAATGAACAATATGAC 804
Oy 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
Db 805 CATGCTTTATGAAAGAGATGCGCAGAGAACATCAAGTTAACCCAGAGATGCCAGTCTCCA 864
Oy 1057 AspAspAlaLysMetAsnGluLysLeuTrpThrValLysAspValAlaMetAsnIleIle 1076
Db 865 GATGATCTCAAGACAAATGAAAAACTGTATACAGATGATGATGCTCTCTGTGTATA 924
Oy 1077 MetSerLysSerThrThrLysSerLeuLysProLysAspProValLeuProAlaArg 1096
Db 925 AATAGTAAAGTCTTGT--GCAATGCAAGATTCCTCAAGGCCCTCCATGAAT 982
Oy 1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTrpLeuProProGluMet 1116
Db 983 TTTTACACA---CCGAAAGAGACTCTGTAAACGATTAAGAGTTATTTTACAGACAGACA 1039
Oy 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
Db 1040 AGAGTACTTCTGTTAAACAGAAAGCCAGCAAGCTGAGTACTGACAGTAAATATAG 1099
Oy 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGlnThrValSer 1156
Db 1100 CTTTATATCAGCAAGCGAAGAGAACCCCTATGTTAAACACACTGAGCAGTGAAGTGAAG 1159
Oy 1157 -----AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg--- 1173

```

```

Db 1160 AATTATTAAATGTAAATTCAGAGCTGAACCTTCAACCGGAATGATCAAGGACAGAGT 1219
Oy 1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp-----TyrThrMetSerSer 1191
Db 1220 TCAGAGCGCAGCAAACTGAGTGTGTAATAATGAAGAACCTGTGAGATTTATTCA 1279
Oy 1192 ProLeuProGlyLysLysSerAspLysTrpAspAspSerAspLeuValArgSerGluLeu 1211
Db 1280 GTCACACCTGTAAAGAAATATTGACCCAGTAAAGAAATAG-----GTAAATAATGCAATTG 1333
Oy 1212 GluLysProArgGlyArgLysLysThrProValThrGlnGlnGlu-GluLysLeuGlyLys 1231
Db 1334 CAAGGAGGAAATGAAAGCCCAACAGAAAGCAGGCTCCAGCTTCGCAAAACTTGAT 1393
Oy 1231 tAspAspLeu-----ThrLysLeuValGlnGln-----Lys 1242
Db 1394 CACAATATGCTGAAACAGAAATGAGCTCACTTCAGAACACACACTCTGCTTGAA 1453
Oy 1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1262
Db 1454 AACTAAAGAGACTATTAATCTCTTTTCACATGACACACAGTCTCTGATGAAATGTACA 1513
Oy 1262 nGlnTrpProGlnGluLysArgLeuLys 1271
Db 1514 GCAGAAACTCTGAGAGAGAGGCTTAAA 1541

RESULT 12
ABS56684
ID ABS56684 standard; cDNA; 1104 BP.
XX
AC ABS56684;
XX
DT 23-JAN-2003 (first entry)
XX
XX
DE Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.
XX
KW Euchromosome fragile intelligence delay syndrome protein 20.02; human;
XX
XX
OS Homo sapiens.
XX
OS
XX
FH
XX
FT CDS location/Qualifiers
FT /*tag= a
FT /*product= "euchromosome fragile intelligence delay
FT syndrome-associated protein, 20.02"
XX
PN
XX
XX
CN1351041-A.
XX
PD 29-MAY-2002.
XX
PF 26-OCT-2000; 2000CN-0125797.
XX
PR 26-OCT-2000; 2000CN-0125797.
XX
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
PI
XX
DR WPI: 2002-637117/69.
XX
DR P-PSDB; ABB84560.
XX
PT Human euchromosome fragile intelligence delay syndrome associated
XX
XX protein 20.02 polypeptide, used to treat e.g. dementia -
XX
PS Claim 6; Page 29-30 (disclosure); 33p; Chinese.
XX
CC This invention describes the novel human euchromosome fragile
XX intelligence delay syndrome associated protein, 20.02. The polypeptide is
XX used in treating diseases such as euchromosome fragile intelligence delay
XX syndrome, and dementia. This sequence encodes the human euchromosome
XX fragile intelligence delay syndrome-associated protein, 20.02 described

```

CC In the disclosure of the invention.

XX Sequence 1104 BP; 362 A; 188 C; 228 G; 326 T; 0 other;

Alignment Scores:

Pred. No.:	3,13e-58	Length:	1104
Score:	967.00	Matches:	183
Percent Similarity:	90.56%	Conservative:	28
Best Local Similarity:	78.54%	Mismatches:	22
Query Match:	13.44%	Indels:	1
DB:	24	Gaps:	0

US-09-512-581B-2 (1-1391) x ABB56684 (1-1104)

```
QY 348 AAlaLysAspLeuThrgLurYrLeuLysValArgSerHisAspProGluGluAlaLeuArg 367
Db 2 GCGAAGATCTCACAGAAATATTTAAAGTTAGATCATCATATCCAGAAAGAGCATTCGT 61
QY 368 HisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp 387
Db 62 CAGGATGTCATTGTTACTATATATACAGCTGCCAGGAGGAGCCTGCC-TTAGATAATGAT 120
QY 388 HisLeuAsnBheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGlu 407
Db 121 CACCTGCTTGCTTTGTGAAGGGAAGACAGCTGATTAACGGTGGCGAGTAAGAAAGAA 180
QY 408 AlaMetMetGlyLeuAlaGlnIleYrYrLysLysTyrrAlaLeuGlnSerAlaAlaGlyLys 427
Db 181 GCTATGATGGCTGCTGGCTGCTTTATAGAAATACGTTTCATGTTGTAAGCAGGAAG 240
QY 428 AspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleYrYrGlnAsn 447
Db 241 GAAGCTGCAGAGAAAGACACCTGATTAAGACAACTTCGATATTTATATACAGAAC 300
QY 448 SerIleAspAspArgLeuLeuValGluArgIlePheHisIleGlnIleValAlaProHisAsn 467
Db 301 AGCATTGACACAAATTTGTTGTAAGAAATCTTGCTGATGATCTTGTCCTCCACACAC 360
QY 468 LeuGluThrThrGluArgMetLysCysLeuYrTyrrLeuYrAlaThrLeuAspLeuAsn 487
Db 361 CTGGAACAGACAGAGAAATGAATGCTTATATTACTTATATGCTATGTTGGATCCAAAT 420
QY 488 AlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLys 507
Db 421 GCGTAAACCTCTCAACGAAATGTGAAAGTGCAGAACATGCTTCGACCACTGATCGC 480
QY 508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
Db 481 GAACATATGATTTGCCAAGACAGCTACATCAAGAGCTACACTGTTCTGCCATGTTTGA 540
QY 528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
Db 541 AAATGATGACACAGCAAGAAATTTGCTGACCCCGGAAGACACAAATTTTGTGAAG 600
QY 548 LysPheThrGlnValLeuGluAspArgLysIleArgLysGlnLeuGluValLeuVal 567
Db 601 AAATTAACCAAGCTTCGCGCATGATGAGAAACTTCGCTCAGATTGGAGTTATATT 660
QY 568 SerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
Db 661 AGCCCAACTGTTCTTGCAACACAGCATATTGTGTG 699
```

RESULT 13

AAK85948 standard; DNA: 7943 BP.

AAK85948;

07-NOV-2001 (first entry)

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.

Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

```
KW cytosolic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
```


PR 25-SEP-2000; 2000US-0234988.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0250393.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure: SEQ ID NO 40760; 3071bp + Sequence Listing: English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK4703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7943 BP: 2567 A; 1175 C; 1468 G; 2733 T; 0 other;

Alignment Scores:
Pred. No.: 2,66-53 Length: 7943
Score: 912,50 Matches: 180
Percent Similarity: 85,31% Conservative: 0
Best Local Similarity: 85,31% Mismatches: 0
Query Match: 12,69% Indels: 31
De: 22 Gaps: 1

US-09-512-581b-2 (1-1391) x AAK85948 (1-7943)

QY 1212 GtutysProArqgGlyArqLysLysThrProValThrgLugIngluLysLugLysMet 1231
Db 2 GAGAAAGCCTAGAGGCGAGAAAAAAGCCCGCCACAGAAACGAGAGAAATTAAGTATG 61

QY 1232 AspAspLeuThrLysLysValGIngluInlInsProLysGlySerGlnArgSerArgLys 1251
Db 62 GATGACTTGACTAACTTGCTGACAGAACAGAACTAAAGGACAGTCAGCAAGTGGGAAA 121

QY 1252 ArgGlyHisThrAlaSerGluSerAspGluGlnGlnIntPrProGluGluLysArgLeuLys 1271
Db 122 AGAGCCCATACGCGCTTCAGAAATCTGATGAACAGCAGTGGCTGAGGAAAAAGGCTCAA 181

QY 1272 GluAspIleLeuGluAsnGluAspGluInlInsnsrPrProLysGlyLysArgGly 1291
Db 182 GAAATATATTAAGAAATGAAGATGAACAGATATGTCGCCAAAAAAGGTAAGAGAGGC 241

QY 1292 ArgProPolysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
Db 242 CGACACCAAAAACCTCTTGTTGGAGGTACACCAAAAAGAGCCAAATGAATGAATCTCT 301

QY 1312 LysLysGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGlu 1331
Db 302 AAAAAAGAGACCAAAAAAATCTGACCTCCAGACCAAGAGAGAGAGAGAGAGAGAGAA 361

QY 1332 ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArg 1351
Db 362 AGACAAAGTGAATATACGAGACAGAGATCCAAAGAACACACACCGAGTGTCAAGAGGA 421

QY 1352 AlaGlnGln----- 1354

D 422 GCACAGCAGAGTACGATGTACTTAACTGCTGTTTCTACTATATTTAA 481
Q 1355 -----ArgalagIuserProglu 1360
D 482 ATCAATAATTGATGCTATCCACATTTGGTCTTCCCAAGACAGAGATCTCCGAA 541
Q 1361 SerSerAlaIleGluSerThrInGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
D 542 TCTAGTGCATTTGAAATCCACACAGCTCCACACAGAAAGAGCGAGAGACCATCAAAA 601
Q 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
D 602 ACGCATCACCATTCACACCAAAAAAATGTG 634
RESULT 14
AAS44727
ID AAS44727 standard; DNA: 714 BP.
AC AAS44727;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #152.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytoskeletal; antirheumatic; antithrilitic; vulnery; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiallergic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN W0200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WC-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Dermanac R;
XX
DR WPI: 2001-589862/66.
DR P-PSDB; AAU27827.
XX
XX Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
XX Claim 1; SEQ ID No 152; 153bp; English.
XX
XX Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a
CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC such as leukemia, lymphoma and neuroblastoma, autoimmune disorders such
CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system

CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).
XX
SQ Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;
XX
Alignment Scores:
Pred. No.: 2.68e-51 Length: 714
Score: 864.00 Matches: 166
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.01% Indels: 0
DB: 22 Gaps: 0
US-09-512-581B-2 (1-1391) x AAS44727 (1-714)
Q 1226 GlnGluLysLeuGlyMetLaspPleuThrLysLeuValGlnGlnGlnLysProLysGly 1245
D 3 GAGGAGAAATTTAGTATGATGACTTGTACTTAAGTGGTACAGAAACGAAATCAAAAGGC 62
Q 1246 SerGlnArgSerArgLysArgGlyHisThrAlaSerGlnSerLysArgGlnGlnIntPro 1265
D 63 AGTCAGCAGAGTCCGAAAGAGCGCATACGCTTCAGAAATCTGATGAACGACGATGGCCT 122
Q 1266 GlnGluLysArgLeuLysGlnAspIleLeuGlnLysGlnLysGlnAsnSerProPro 1285
D 123 GAGGAAAGAGGCTCCAAAGAGATATTTAGAAATGCAATGCAAGCAATAGTCCGCCA 182
Q 1286 LysLysGlyLysArgGlyArgProProLysProLeuGlyGlyThrProLysGlnGlu 1305
D 183 AAAAAGGCTAAAGAGGCCCAACCAACCAAACTCTTGCTGAGGTACCAACCAAGAGAG 242
Q 1306 ProThrMetLysThrSerLysLysGlySerLysLysSerGlyProProAlaProGlu 1325
D 243 CCAACATTTGAAACTTCTTAATAAAGAGAAAGCAAAAAAATCTGACCTCCAGCAGAG 302
Q 1326 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1345
D 303 GAGGAGGAAAGAGAAAGAAAGCAAAATGCAATACGGAACGAAAGTCCAAAAACCAACAG 362
Q 1346 HisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlnSerSerAlaIleGlu 1365
D 363 CACCGAGTGTCCAAAGAGAGACACACAGACAGAGCAATCTCTTAATCTAGTGCATTTGAA 422
Q 1366 SerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer 1385
D 423 TCCACAGATGCCACACACAGAAAGAGCGAGAAAGAACATCAAAACGCATCATCATCA 482
Q 1386 GlnProLysLysAsnVal 1391
D 483 CAACCAAAAAAATGTG 500
RESULT 15
AAF18338
ID AAF18338 standard; DNA: 1217 BP.
XX
AC AAF18338;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 357.
XX

KM Human; lung cancer associated protein; neuroprotective; cytostatic;
 KM cardiotoxic; immunomodulatory; muscular active; vulnerary;
 KM gastrointestinal; nephrotoxic; antineoplastic; gynecological;
 KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KM proliferative disorder; wound healing; infectious disease; ds.
 OS Homo sapiens.
 XX MO20005180-A2.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05918.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Ruben SM;
 DR WPI: 2000-587514/55.
 DR P-PSDB: AAB58462.
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 PS Claim 1; Page 815; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1217 BP; 445 A; 232 C; 270 G; 263 T; 7 other:
 Alignment Scores:
 Pred. No.: 6,38e-44 Length: 1217
 Score: 762.50 Matches: 185
 Percent Similarity: 56.88% Conservative: 63
 Best Local Similarity: 42.43% Mismatches: 121
 Query Match: 10.60% Indels: 68
 DB: 21 Gaps: 15
 US-09-512-581B-2 (1-1391) x AAF18338 (1-1217)
 QY 965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaIleValSerGluLysLeuLeu 984
 DB 2 AAFATACGATATACGAGGATACATTAACGAGATCTTATGCGTACGAGAAATTTATTA 61
 QY 985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
 DB 62 TCACGTGTGCTGATATATGATGTTCCATACATGATTCACCTGCTGAGCCCATGATCCAGAT 121
 QY 1005 TyrValLysValGlnAspIleGluLeuLysAspValLysGluCysLeuTyrPheVal 1024
 DB 122 TTTCACAGATCACAGATGTTGATCAGCTGCTGATATCAAGAGTGCCTATGTTGTCATG 181

QY 1025 LeuGluIleuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
 DB 182 CTTGACAGTTTAAATGACAAAGATGAAATGAAATGACCATGCTTTATGACAAAGATGCA 241
 QY 1045 GluAsnIleLysGlnThrLysAspAlaGlnGluProAspAspAlaLysMetAsnGluLys 1064
 DB 242 GAGAACATCAAGTTAAACGAGATGCCAGTCTCCAGATGATCCACAAACAAATGCAAAA 301
 QY 1065 LeuTyrThrValLysAspValAlaMetLysIleIleMetSerLysSerThrThrTyrSer 1084
 DB 302 CTGATACAGTATGATGATGCTGCTGCTGTTAAATGATGAAAGTCTTTGTGCAAT 361
 QY 1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
 DB 362 GCAGATTCACCAAGGAGGAGCCAGTCTCCCATGAAATTTTATACACACGCAAAAGAC 421
 QY 1105 PheSerAsnThrLysAsnTyrLeuProGluMetLysSerPhePheThrProGlyLys 1124
 DB 422 TTCTGTAAAGATTAATATATTTACAGAAAGACATCTTCTTAAACAGAAAG 481
 QY 1125 ProLysThrThrAsnValLeuGluValAlaLysLysProLysSerSerAlaGlyLysGln 1144
 DB 482 CCAAGGCTGCTGGAGTACTGAGTGCAGTAAATTAACCTTTATACACGAGGAAAGAAA 541
 QY 1145 SerGlnThrLysSerSerArgMetGluThrValSer-----AsnAlaSerSerSer 1162
 DB 542 CCTATGTTAAGACACATGCGACATGACCTGGAACATATTAATTAATTCAGAGCTG 601
 QY 1163 AsnProSerSerProGlyArgIleLysGlyArg--LeuAspSerSerGluMetAspHis 1181
 DB 602 AACCTTCAACCGGAAATCGATCAAGGAAACGAGTTCAGAGGACAGAAATCGAGATT 661
 QY 1182 SerGluAsnGluAsp-----TyrThrMetSerSerProLeuProGlyLysLysSerAsp 1199
 DB 662 ACTGAAATGAGAGAACCCCTGTCAGCATTTTTCAGTCACACCTGTTAAACAAATTTGAC 721
 QY 1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys 1219
 DB 721 ----- 721
 QY 1220 ThrProValThrGluGlnGluGluLysLeuGluMetAspAspLeuThrLysLeuValGln 1239
 DB 722 --CCAGTAAAGAAATTAAGAA-----ATTAAATCTGATCGGCTAAC----- 760
 QY 1240 GluGluLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
 DB 761 ---CAGGCAACATCAGCAGTGCAGCGAGAAAGAAAGAAACAGTAACAGCGCTGTGCA 817
 QY 1260 AspGlu---GlnGlnTyrProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278
 DB 818 GAGATATTCACAAACAAACAAAGATGAGAAA-----CTA 850
 QY 1279 AspGluGlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeu 1297
 DB 851 GATGAAATGGGAGCTCCCGCCCTCCAAACCCAGAGAGAGAGCTGCACCAACATCTGAA 910
 QY 1298 Gly---GlyGlyThrProLysGluGluProThrMetLysThrSerLysGlySerLys 1316
 DB 911 TTCAGGGCAATGCTACCAAAATGATGATCTTAATTAACCTATTAAACAGGAGAGAG 970
 QY 1317 LysLysSer---GlyProProAlaProGluGluGlu----- 1338
 DB 971 AGAGCTGAGTGGGTACAGAGAGAGCTGGGGTGTGAGAGAGGTAATGCCAAAGACACC 1030
 QY 1339 -----GluGluGluArgGlnSerGlyAsnThrGlnLys-----SerLys 1342
 DB 1031 AAATGCAAGATTT-AGCCAAAAAGGCGACACGACGAAAGACMAATGACTTACAAAG 1089
 QY 1343 SerLysGlnHis-----ArgValSerArgArgLaglnArgLaglnArgLysProGlu 1360
 DB 1090 GTRAAATGCAATTTGCAAAAGGAGGAAATGAAGGCCAAACGAAAGAGCTTCCAGTTT 1149
 QY 1361 SerSerAlaIle-----GluSerThrGlnSerThr 1370

Db 1150 TGC AAAA NCTTGGATTACAATGCGCTGACAGAAATGACTTATTCAACC 1197

Search completed: September 24, 2003, 20:46:25
Job time : 949 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 19:50:39 : Search time 175 seconds
(without alignments)
2051.150 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRTNDGKITPPGVK.....QKGRGPKSPSPQPKKNV 1391

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7155	99.5	1447	4	Q9NT15
2	7155	99.5	1483	4	Q9Y215
3	7101	98.7	1391	4	Q9Y451
4	4668.5	64.9	1297	4	Q8NG14
5	3787	52.6	824	11	Q8BLH6
6	3456	48.0	701	11	Q8BJ18
7	2698.5	37.5	851	4	Q9Y4D4
8	2517	35.0	529	4	Q8IXT6
9	2487	34.6	479	4	Q96KV4
10	2476.5	34.4	600	4	Q96KV3
11	2474	34.4	600	4	Q8N7J4
12	2317	32.2	584	11	Q8VDS0
13	1982.5	27.6	1218	5	Q9V6A6
14	833.5	11.6	1205	3	Q9HFF5
15	808	11.2	1638	10	Q9FTL0
16	807	11.2	1579	5	Q9XTF6

17	791	11.0	198	11	Q9D337	Q9D337 mus musculus
18	747.5	10.4	1596	3	Q9UYV6	Q9UYV6 sordaria ma
19	656	9.1	125	4	Q9H5N8	Q9H5N8 homo sapien
20	640.5	8.9	333	4	Q96DB6	Q96DB6 homo sapien
21	632.5	8.8	1506	3	Q94076	Q94076 emericella
22	602	8.4	192	11	Q8BX77	Q8BX77 mus musculus
23	572.5	8.0	1378	10	Q8S1N0	Q8S1N0 oryza sativ
24	571.5	7.9	1303	10	Q9CAP7	Q9CAP7 arabidopsis
25	487	6.8	415	5	Q9SU25	Q9SU25 drosophila
26	423.5	5.9	1033	5	Q8MWY4	Q8MWY4 dictyostell
27	309.5	4.3	390	3	Q94237	Q94237 schizosacch
28	303	4.2	780	10	Q9SA10	Q9SA10 arabidopsis
29	264	3.7	990	10	Q9SP90	Q9SP90 arabidopsis
30	256	3.6	873	10	Q8GUP3	Q8GUP3 arabidopsis
31	248.5	3.5	2158	4	Q9Y673	Q9Y673 homo sapien
32	242.5	3.4	2265	4	Q9Y674	Q9Y674 homo sapien
33	235	3.3	1135	6	Q8H2M3	Q8H2M3 canis fam11
34	234.5	3.3	852	10	Q9SZ55	Q9SZ55 arabidopsis
35	234.5	3.3	3130	5	Q8IDX6	Q8IDX6 plasmodium
36	232.5	3.2	3130	5	Q9BK46	Q9BK46 plasmodium
37	231.5	3.2	2253	13	P70012	P70012 xenopus lae
38	230	3.2	2501	5	Q9NCW7	Q9NCW7 drosophila
39	229	3.2	1723	2	Q9JMX8	Q9JMX8 helicobacte
40	228.5	3.2	18074	5	Q917U4	Q917U4 drosophila
41	225	3.1	1437	5	Q18033	Q18033 caenorhabdi
42	223.5	3.1	4717	3	Q94248	Q94248 schizosacch
43	223	3.1	2139	5	Q07569	Q07569 entamoeba h
44	223	3.1	3254	5	Q9BK45	Q9BK45 plasmodium
45	222.5	3.1	1819	16	Q9ZLV0	Q9ZLV0 helicobacte

ALIGNMENTS

RESULT 1

ID Q9NT15 PRELIMINARY: PRT: 1447 AA.
AC Q9NT15:
DT 01-OCT-2000 (TREMBLrel. 15. Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP
SEQUENCE FROM N.A.
RA Rhodes S., Huckle E.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL137201; CAB69911.1; -
DR InterPro, IPR000637; AT_hook.
DR SMART, SM00384; AT_hook; 1.
KW Hypothetical protein.
SQ
SEQUENCE 1447 AA; 16466 MW; 145C30308EA3EFD5 CRC64;

Query Match	99.5%	Score 7155:	DB 4:	Length 1447:
Best local similarity	99.7%	Pred. No. 0:		
Matches 1387:	Conservative 0:	Mismatches 4:	Indels 0:	Gaps 0:
QY	1	MAHSKTRTNDGKITPPGVKESIDKISKEEWRLKVVKTFMDODDSEEEKELYNTLA	60	
DB	1	MAHSKTRTNDGKITPPGVKESIDKISKEEWRLKVVKTFMDODDSEEEKELYNTLA	60	
QY	61	LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKDIFMETRQLKGLDPT	120	
DB	61	LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKDIFMETRQLKGLDPT	120	
QY	121	KSPQFNRYEVLLENIAWKSYNICELEDSNEITOLYRTLFYINNGHONKVMHWDL	180	
DB	121	KSPQFNRYEVLLENIAWKSYNICELEDSNEITOLYRTLFYINNGHONKVMHWDL	180	
QY	181	MSITICEDGVSOELDPTVLVNLVPAHKNLKQAYDLAKALLKRTAQAIEPYITTFPNV	240	

```

|||||
Db 181 MSIIIEGDTVSQELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFQV 240
Qy 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPQLEFKLSNDNEERQVYKLLAKMF 300
Db 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPQLEFKLSNDNEERQVYKLLAKMF 300
Qy 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKRS 360
Db 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKRS 360
Qy 361 DPEEAI RHVDYIVSYTAAKKDILLVNDHLNFEVREBTLDRMVRKREAMGLAQYKKYA 420
Db 361 DPEEAI RHVDYIVSYTAAKKDILLVNDHLNFEVREBTLDRMVRKREAMGLAQYKKYA 420
Qy 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVPHNLETTERMKCLYYL 480
Db 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVPHNLETTERMKCLYYL 480
Qy 481 YATLDINAVKALNEMKCONLHRQVLDLILKOPTDASVKAIFSKVAVITRNLPDPG 540
Db 481 YATLDINAVKALNEMKCONLHRQVLDLILKOPTDASVKAIFSKVAVITRNLPDPG 540
Qy 541 KAODFMKKFTQVLEDEKIRKOLEVLVSPTCSCKQAEQVREITKKLGPNKOPTNPLEM 600
Db 541 KAODFMKKFTQVLEDEKIRKOLEVLVSPTCSCKQAEQVREITKKLGPNKOPTNPLEM 600
Qy 601 IKRLLEIARVNIHDTESISALIKOVNKSIDGTADDEGEVPTQOATIRAGIELLKVSFT 660
Db 601 IKRLLEIARVNIHDTESISALIKOVNKSIDGTADDEGEVPTQOATIRAGIELLKVSFT 660
Qy 661 PISFSAEFESLFIACIKMDEKVAEALQIFKNFGSKIEDEPHIRISALLPYLHNHKK 720
Db 661 PISFSAEFESLFIACIKMDEKVAEALQIFKNFGSKIEDEPHIRISALLPYLHNHKK 720
Qy 721 GPRROAKYAIHCHIAIFSSKETQFAQIFEPRLHKSLEPSNIEHLITPLVITGHIALAPDQ 780
Db 721 GPRROAKYAIHCHIAIFSSKETQFAQIFEPRLHKSLEPSNIEHLITPLVITGHIALAPDQ 780
Qy 781 FAAPWKSVAATFIVKDLMDMDRLPGKTKTLMWDEPVSPEVAVKIOATIMVIRWLLG 840
Db 781 FAAPWKSVAATFIVKDLMDMDRLPGKTKTLMWDEPVSPEVAVKIOATIMVIRWLLG 840
Qy 841 NNHSGSGTSLRLTLILHSDGDLTEQGISKPDMSRLRLAAGSALYKLAQEPCEYHEIT 900
Db 841 NNHSGSGTSLRLTLILHSDGDLTEQGISKPDMSRLRLAAGSALYKLAQEPCEYHEIT 900
Qy 901 LEOYQCALAIINDECTQOVQVFAOKLHKLISRLPLEYATCALCAKDPYKERRAHARQ 960
Db 901 LEOYQCALAIINDECTQOVQVFAOKLHKLISRLPLEYATCALCAKDPYKERRAHARQ 960
Qy 961 CLVKNINVRREYIKONAAVSEKLLSLPEYVVPYTHLHAHPDYVVOIDOLKDYKEC 1020
Db 961 CLVKNINVRREYIKONAAVSEKLLSLPEYVVPYTHLHAHPDYVVOIDOLKDYKEC 1020
Qy 1021 LMFVLETLAKNENNSHAFIRKVENIKQTKDAQOPDAAKNEKLYTVCVANMIINSKS 1080
Db 1021 LMFVLETLAKNENNSHAFIRKVENIKQTKDAQOPDAAKNEKLYTVCVANMIINSKS 1080
Qy 1081 TTSLSLSPKDPVLPARFTQPDKNFSNTKNYLPPEMKSFTPGPKTTNVLGAANKPLSS 1140
Db 1081 TTSLSLSPKDPVLPARFTQPDKNFSNTKNYLPPEMKSFTPGPKTTNVLGAANKPLSS 1140
Qy 1141 AGKOSQTSRMETVNASSSNPSPGRIFKGRDSEMDHSEMEDYTMSSPLPGKSKDK 1200
Db 1141 AGKOSQTSRMETVNASSSNPSPGRIFKGRDSEMDHSEMEDYTMSSPLPGKSKDK 1200
Qy 1201 RDDSIVRSELEKPRGRKKTPTVEQEBKLGMDLTKLVQOBKPGSGORSKRKGTASD 1260
Db 1201 RDDSIVRSELEKPRGRKKTPTVEQEBKLGMDLTKLVQOBKPGSGORSKRKGTASD 1260
Qy 1261 EQOWPEKRLKEDILENEDEONSPPKKGRRGPRPKPLGGGTPKKEEPLMKTSKSGSKSG 1320
Db 1261 EQOWPEKRLKEDILENEDEONSPPKKGRRGPRPKPLGGGTPKKEEPLMKTSKSGSKSG 1320

```

```

Db 1261 EQOWPEKRLKEDILENEDEONSPPKKGRRGPRPKPLGGGTPKKEEPLMKTSKSGSKSG 1320
Qy 1321 PAPAEEREEEROSGTEQKSKSKORHVRRAQOARESPSSAISTOSTPOKGRGPRSK 1380
Db 1321 PAPAEEREEEROSGTEQKSKSKORHVRRAQOARESPSSAISTOSTPOKGRGPRSK 1380
Qy 1381 TSPSPQPKKNV 1391
Db 1381 TSPSPQPKKNV 1391

RESULT 2
Q9Y215
ID Q9Y215 PRELIMINARY; PRT: 1483 AA.
AC Q9Y215;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE KIAA0979 protein (Fragment).
GN KIAA0979.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023196; BAA76823.2; -
DR InterPro: IPR000637; AT_Book.
DR PRINTS: PR00929; ATBOOK.
FT NON_TER
SQ SEQUENCE 1483 AA; 168415 MW; 0B328407FD0BDD74 CRC64;

Query Match 99.5%; Score 7155; DB 4; Length 1483;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1387; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MASHKRTNDGKITYPGVKEISDKISKEEMVRLKVVYTFMDMDQDSEEEKELYLNTA 60
Db 37 MASHKRTNDGKITYPGVKEISDKISKEEMVRLKVVYTFMDMDQDSEEEKELYLNTA 96
Qy 61 LHLASDFELKHPKDVRLVLAACCLADIFRIYAEARYTSPDKLIDIFMETTRQLKGLDPT 120
Db 97 LHLASDFELKHPKDVRLVLAACCLADIFRIYAEARYTSPDKLIDIFMETTRQLKGLDPT 156
Qy 121 KSPQFNRYFYLEINIMVSYNICEFLEDSNEIFQTQYRFLSVINNGHOKVHMHVDL 180
Db 157 KSPQFNRYFYLEINIMVSYNICEFLEDSNEIFQTQYRFLSVINNGHOKVHMHVDL 216
Qy 181 MSIIIEGDTVSQELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFQV 240
Db 217 MSIIIEGDTVSQELDTVLVNLVPAHKNLNKOAYDLAKALLKRTQAITEPYITTFNFQV 276
Qy 241 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPQLEFKLSNDNEERQVYKLLAKMF 300
Db 277 LMGKTSISDLSHVEDLILELYNIDSHLLSVLPQLEFKLSNDNEERQVYKLLAKMF 336
Qy 301 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKRS 360
Db 337 GAKDSELASQNKPLMOCYIGRFNDIHVPIRLBECVKAASHCLMNHPLAKDLTEYLVKRS 396
Qy 361 DPEEAI RHVDYIVSYTAAKKDILLVNDHLNFEVREBTLDRMVRKREAMGLAQYKKYA 420
Db 397 DPEEAI RHVDYIVSYTAAKKDILLVNDHLNFEVREBTLDRMVRKREAMGLAQYKKYA 456
Qy 421 LOSAAGKDAKQIAWIKDKLILHYONSIDRLIVERIFAQYVPHNLETTERMKCLYYL 480

```

```

Db 457 LOSAAGKDAKQJAWIKDKLHITYQNSIDDLVERIFRQVWVPIINLETTEKMLXYL 516
QY 481 YATLDLNAVKALNEMKCONLLRHQVKDLDLIKOPKTDASVKAIFSKVWVITRNLPDGC 540
Db 517 YATLDLNAVKALNEMKCONLLRHQVKDLDLIKOPKTDASVKAIFSKVWVITRNLPDGC 576
QY 541 KADPFMKFTQVLEDEDEKIRKOLEVLVSPSCCKQAGCVREITTKLGNPKOTNPFLEM 600
Db 577 KADPFMKFTQVLEDEDEKIRKOLEVLVSPSCCKQAGCVREITTKLGNPKOTNPFLEM 636
QY 601 IKFLERLAPVHIDFESISALIKQVKSIDGTADDEGVPTDQARAGLELLKVLSTFH 660
Db 637 IKFLERLAPVHIDFESISALIKQVKSIDGTADDEGVPTDQARAGLELLKVLSTFH 696
QY 661 PLSFHSAEFPFESLACKMDKDEKVAEALQIFKNKSGKIEDEFPHARSALLPVLLHKKSK 720
Db 697 PLSFHSAEFPFESLACKMDKDEKVAEALQIFKNKSGKIEDEFPHARSALLPVLLHKKSK 756
QY 721 GPPROAKVAIHCIAHAFSSKETQFAQIFEPRLHKSIDPSNLEHITPVLVIGHIALLAPDQ 780
Db 757 GPPROAKVAIHCIAHAFSSKETQFAQIFEPRLHKSIDPSNLEHITPVLVIGHIALLAPDQ 816
QY 781 FAAPKWSVATFIVKDLNMDRLPGKKTTLVWPDDEVSPETVWKIQATIKMVRWLLGK 840
Db 817 FAAPKWSVATFIVKDLNMDRLPGKKTTLVWPDDEVSPETVWKIQATIKMVRWLLGK 876
QY 841 NNHSGSGTSLRLTLTILHSDGDLTBOGKISKPDMSRLAAGSAIVKAIOEPCYHEIT 900
Db 877 NNHSGSGTSLRLTLTILHSDGDLTBOGKISKPDMSRLAAGSAIVKAIOEPCYHEIT 936
QY 901 LEOYOLCALAIDECYQVQVFAOKLHKLGLSLRLPLLEMAICALCAKPVKERRAHQ 960
Db 937 LEOYOLCALAIDECYQVQVFAOKLHKLGLSLRLPLLEMAICALCAKPVKERRAHQ 996
QY 961 CLVKNINVRREYLKQHAASEKLSILPEYVVPYTIHLLAHDPDYKODIEQLKDYKEC 1020
Db 997 CLVKNINVRREYLKQHAASEKLSILPEYVVPYTIHLLAHDPDYKODIEQLKDYKEC 1056
QY 1021 LMFVLEILMAKNENSHAFIRKVENIKOTKDAQGDDAKMNEKLTVCDAVANNITMSK 1080
Db 1057 LMFVLEILMAKNENSHAFIRKVENIKOTKDAQGDDAKMNEKLTVCDAVANNITMSK 1116
QY 1081 TTYTLESPPKDPVLPARFTQPDKNFSNTKNYLPRPEKSFPTPKPTTNVLGAVNKLSS 1140
Db 1117 TTYTLESPPKDPVLPARFTQPDKNFSNTKNYLPRPEKSFPTPKPTTNVLGAVNKLSS 1176
QY 1141 AKQOSOTKSSRMETVSNASSSSNPSSPGRIKGRLDSEMDHSEMDYTMSPLPKKSDK 1200
Db 1177 AKQOSOTKSSRMETVSNASSSSNPSSPGRIKGRLDSEMDHSEMDYTMSPLPKKSDK 1236
QY 1201 RDDSLVRSLELKPGRKKTPTTEOEKLGMDLTTLVQEQPKGQSRKRGHTASED 1260
Db 1237 RDDSLVRSLELKPGRKKTPTTEOEKLGMDLTTLVQEQPKGQSRKRGHTASED 1296
QY 1261 EQOWPEEKRLKEDILENDEQNSPPKKGKRGPRPKPLGGTPEKEPTMTSKKSGKSG 1320
Db 1297 EQOWPEEKRLKEDILENDEQNSPPKKGKRGPRPKPLGGTPEKEPTMTSKKSGKSG 1356
QY 1321 PPAPEEEREEERQSGTEOKSKQHRVSRRAQRAESPESAIESTOSTPOKGRGRPSK 1380
Db 1357 PPAPEEEREEERQSGTEOKSKQHRVSRRAQRAESPESAIESTOSTPOKGRGRPSK 1416
QY 1381 TPSPSOPKKNV 1391
Db 1417 TPSPSOPKKNV 1427

```

RESULT 3
 09Y451 PRELIMINARY: PRT: 1391 AA.
 AC 09Y451:
 DT 01-NOV-1999 (TReMBLrel. 12, Created)

```

DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE Androgen-induced prostate proliferative shut-off associated
DE protein.
GN AS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
RT Tissue=Prostate;
RT "Androgen induced proliferative shut-off in prostate cancer cells.";
RL Proc. Am. Assoc. Cancer Res 37:223-223(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
RT Tissue=Prostate;
RT "Early gene expression during androgen-induced inhibition of
RT proliferation of prostate cancer cells: a new suppressor candidate on
RT chromosome 13, in the BRCA2-Rb1 locus.";
RL J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
DR EMBL: U95825; AAD2134.2;
DR InterPro: IPR000637; AT_hook.
DR SMART: SM00384; AT_hook; 1.
SQ SEQUENCE 1391 AA; 158035 MW; F58AEES5AD6D9479 CRC64;

Query Match 98.7%; Score 7101; DB 4; Length 1391;
Best local similarity 99.0%; Pred. No. 0;
Matches 1377; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 MAHSTRINDGKITTPPGVKESIDKISKEMVRLKAVVKTMDODSEEEKELYLNA 60
Db 1 MAHSTRINDGKITTPPGVKESIDKISKEMVRLKAVVKTMDODSEEEKELYLNA 60
QY 61 LHLASDFLKHGKDVRLVACCLADIFRIVAPARYSPDKIOIFMTQGLGLED 120
Db 61 LHLASDFLKHGKDVRLVACCLADIFRIVAPARYSPDKIOIFMTQGLGLED 120
QY 121 KSPQNRVRYLLLENIAWVSYNICELEDSNEIFQLYRTLSVINGNHNKVMHMYDL 180
Db 121 KSPQNRVRYLLLENIAWVSYNICELEDSNEIFQLYRTLSVINGNHNKVMHMYDL 180
QY 181 MSIIICEGTVSOELDTVLVNLVPAHKNLKNQAVDLAKALKKRTAQAIEPTTFNFQV 240
Db 181 MSIIICEGTVSOELDTVLVNLVPAHKNLKNQAVDLAKALKKRTAQAIEPTTFNFQV 240
QY 241 LMLGTSISDLSSEHFDTLLELYNIDSLLSVLPOLFEKLSNDNEERLQVVKLLAKMF 300
Db 241 LMLGTSISDLSSEHFDTLLELYNIDSLLSVLPOLFEKLSNDNEERLQVVKLLAKMF 300
QY 301 GAKDELSAQNPMLQCYLGRNDIHVPTRLECVFASHCILNHPDLAKDLELYKVRSH 360
Db 301 GAKDELSAQNPMLQCYLGRNDIHVPTRLECVFASHCILNHPDLAKDLELYKVRSH 360
QY 361 DPEEAI RHDIYISYTAACKDILLVNDHLNVREERTLDKPRKRVKKAAMGLAQIYKKA 420
Db 361 DPEEAI RHDIYISYTAACKDILLVNDHLNVREERTLDKPRKRVKKAAMGLAQIYKKA 420
QY 421 LOSAAGKDAKQJAWIKDKLHITYQNSIDDLVERIFRQVWVPIINLETTEKMLXYL 480
Db 421 LOSAAGKDAKQJAWIKDKLHITYQNSIDDLVERIFRQVWVPIINLETTEKMLXYL 480
QY 481 YATLDLNAVKALNEMKCONLLRHQVKDLDLIKOPKTDASVKAIFSKVWVITRNLPDGC 540
Db 481 YATLDLNAVKALNEMKCONLLRHQVKDLDLIKOPKTDASVKAIFSKVWVITRNLPDGC 540
QY 541 KADPFMKFTQVLEDEDEKIRKOLEVLVSPSCCKQAGCVREITTKLGNPKOTNPFLEM 600
Db 541 KADPFMKFTQVLEDEDEKIRKOLEVLVSPSCCKQAGCVREITTKLGNPKOTNPFLEM 600

```

QY 601 IKFLERIAVPHIDTESISALIKOVNKSIDGTADDEGEVPTDQAIRAGLELLKVLSTFH 660
 Db 601 IKFLERIAVPHIDTESISALIKOVNKSIDGTADDEGEVPTDQAIRAGLELLKVLSTFH 660
 QY 661 PISHSMETPESLLACLAKMDEKVAEALQIFKNTGSKIEEDFPHISALLPVJHNSKK 720
 Db 661 PISHSMETPESLLACLAKMDEKVAEALQIFKNTGSKIEEDFPHISALLPVJHNSKK 720
 QY 721 GPPQAKYAIHCHIAIFSSKETQFAQIFEPHLKSLDPSNLEHLTPVTYIGHIALLPDQ 780
 Db 721 GPPQAKYAIHCHIAIFSSKETQFAQIFEPHLKSLDPSNLEHLTPVTYIGHIALLPDQ 780
 QY 781 PAAPWKSVAATFIVKDLMDRLPGKKTTLMPDEEVSPEYMWKIQAIKMMVWLLGKM 840
 Db 781 PAAPWKSVAATFIVKDLMDRLPGKKTTLMPDEEVSPEYMWKIQAIKMMVWLLGKM 840
 QY 841 NNHSGSTSLRLITLTHSDGDLTEQKISKPMMSRLAAGSAIYKLAQEPYHEIT 900
 Db 841 NNHSGSTSLRLITLTHSDGDLTEQKISKPMMSRLAAGSAIYKLAQEPYHEIT 900
 QY 901 LEQYQCALAINDECYQVQFAOKLHKGSLRLPLPEYMAICALCAKDPYKERRAHARQ 960
 Db 901 LEQYQCALAINDECYQVQFAOKLHKGSLRLPLPEYMAICALCAKDPYKERRAHARQ 960
 QY 961 CLVKNINVRREYIKQHAASEKLLSLPEYVVPYTHLLADHPDYKQVODIEQLKDYEC 1020
 Db 961 CLVKNINVRREYIKQHAASEKLLSLPEYVVPYTHLLADHPDYKQVODIEQLKDYEC 1020
 QY 1021 LMFVLEITLMAKNENSHAFTRKVENIKOTKADGPDADAKNEKLYTVCDVAMNITMSKS 1080
 Db 1021 LMFVLEITLMAKNENSHAFTRKVENIKOTKADGPDADAKNEKLYTVCDVAMNITMSKS 1080
 QY 1081 TTVSLESBPKDVLPAFEFTQPKDNFSNTKNLPEPMKSFETPGKPTTNVLGAVNKPLSS 1140
 Db 1081 TTVSLESBPKDVLPAFEFTQPKDNFSNTKNLPEPMKSFETPGKPTTNVLGAVNKPLSS 1140
 QY 1141 AGKQSQTKSSMRETVNASSSSNPSPGRKIGRLDSSBMDHSEMDYTMSSPLGKKSNDK 1200
 Db 1141 AGKQSQTKSSMRETVNASSSSNPSPGRKIGRLDSSBMDHSEMDYTMSSPLGKKSNDK 1200
 QY 1201 RDDSDVLRSELEKPRGRKKTPTVTEQEEKLGMDDLTKLYOEOQPKGSORSRKRGHTASESD 1260
 Db 1201 RDDSDVLRSELEKPRGRKKTPTVTEQEEKLGMDDLTKLYOEOQPKGSORSRKRGHTASESD 1260
 QY 1261 EQQPEEKRLKEDILLENEDEQNSPPKKGKRGPRPKPLGGGTPEDEPTMKTSSKSKSSG 1320
 Db 1261 EQQPEEKRLKEDILLENEDEQNSPPKKGKRGPRPKPLGGGTPEDEPTMKTSSKSKSSG 1320
 QY 1321 PPAPDEEVEEROGNTQKSKSKQHVSRRAQRAESSAIESTOSTPQKGRGRPSK 1380
 Db 1321 PPAPDEEVEEROGNTQKSKSKQHVSRRAQRAESSAIESTOSTPQKGRGRPSK 1380
 QY 1381 TPSPSPQPKKNV 1391
 Db 1381 TPSPSPQPKKNV 1391

RESULT 4

Q8NG14 PRELIMINARY; PRT: 1297 AA.
 AC Q8NG14: 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 GN SCC-112.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RA Kumar D., Patel S., Whiteside T.L., Kasid U.;

RT "Identification and characterization of SCC-112, a novel cell cycle
 RT regulated gene in human cancer cells."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF294791; AAM82347.1; -
 SQ SEQUENCE 1297 AA; 146610 MW; A42D13E68CA508E CRC64;
 Query Match 64.9%; Score 4668.5; DB 4; Length 1297;
 Best Local Similarity 69.0%; Pred. No. 2.1e-261;
 Matches 908; Conservative 160; Mismatches 199; Indels 49; Gaps 11;
 QY 31 MVRRLKVVVTFMDMODSEBEKLYNLALHLASDFLKHGKDVALLVACCAIDFRI 90
 Db 1 MIKRLKVVVTFMDMODSEBEKQOYPLALHLASEFLRNPNKDVLLVACCAIDFRI 60
 QY 91 YAPAPPTSPDKLIDIMEFTIROLKGLDPRSPQNFYFVLENIAMVKSYNICELEDS 150
 Db 61 YAPAPPTSHDKLIDILFITRQLKGLDPRSPQNFYFVLENIAMVKSYNICELEDC 120
 QY 151 NEIFTQYRLFSVYINNGHNOVHMHWMDLMSIICBGDTVSOBLDTVLNIVPAHKNL 210
 Db 121 NEIFTQYRLFSVYINNSHKKVQHMMDLMSIIMGEGVTOELLSIILNIPAHKNL 180
 QY 211 NKQAYDLAKALKTKTAIEPYITTFNOVLMGKTSISDSEVDFLIELYIDSHLL 270
 Db 181 NKQSFDLAKVLLKRTVOTIACLANFNOVLVGRSSVSDSEHVFQDLELFAIDPHL 240
 QY 271 LSVLPQLEFKLSNDNEERLOVYKLMKMGAKSELASONKPLMOCYLGRFNDIHYPIR 330
 Db 241 LSVLPQLEFKLSNDGERLAVVRLAKLFGSKODLATQNRPLMOCYLGRFNDIHYPIR 300
 QY 331 LECVKEFASHCLMNHDLAKDLTEYLVKVRSHDPEEAIHHDVIVSVITAAKDLILVNDHL 390
 Db 301 LESYKFAFASHCLMNHDLAKDLTEYLVKVRSHDPEEAIHHDVIVSVITAAKDLILVNDQL 360
 QY 391 NFVERFLDKRWRKREKAMMGLAOLYKRYKYLQSAAGDAKQALAMIKDLIHYONSID 450
 Db 361 GFVERFLDKRWRKREKAMMGLAOLYKRYKYLQSAAGDAKQALAMIKDLIHYONSID 420
 QY 451 DRLVERIFIAQYVPHNLEETTERKCLYUYATLDAVNAVALNEMKQNLRLHYQVLDL 510
 Db 421 DRLVERIFIAQYVPHNLEETTERKCLYUYATLDAVNAVALNEMKQNLRLHYQVLDL 480
 QY 511 DLKQPKTDAVKAIFSKVAVYTRNLPDQKADQPMKKFVOVLEDEKIRKOLEVLYSPT 570
 Db 481 DLKQPKTSEANCSAMFKLMTIAKNLPDPKADQFVKKFNOVLGDDDKLSOLELLISPT 540
 QY 571 CSCQKQAGCVREIKTKGKOPNPNFLMIKFLERIAVPHIDTESISALIKOVNSID 630
 Db 541 CSCQKQADICVREIARKLANPNKOPNPNFLMIKFLERIAVPHIDTESISALIKOVNSID 600
 QY 631 GTADEDEGVPTDQAIRAGLELLKVLSTFHPISFSAETESLACLAKMDEKVAEALQ 690
 Db 601 GTADEDEGVPTDQAIRAGLELLKVLSTFHPISFSAETESLACLAKMDEKVAEALQ 660
 QY 691 IFKNTGSKIEEDFPHISALLPVJHNSKSGPPQAKYAIHCHIAIFSSKETQFAQIFEP 750
 Db 661 IFKNTGSKIEEDFPHISALLPVJHNSKSGPPQAKYAIHCHIAIFSSKETQFAQIFEP 720
 QY 751 LKHSIDPSNLEHLTPVTYIGHIALLPDQFAAPWKSVAATFIVKDLMDRLPGKKT 810
 Db 721 LKHSIDPSNLEHLTPVTYIGHIALLPDQFAAPWKSVAATFIVKDLMDRLPGKKT 780
 QY 811 LMPDEEVSPEYMWKIQAIKMMVWLLGKMNNHSGSTSLRLITLTHSDGDLTEQGI 870
 Db 781 LMPDEEVSPEYMWKIQAIKMMVWLLGKMNNHSGSTSLRLITLTHSDGDLTEQGI 840
 QY 871 SKPMSRLRLAAGSAIYKLAQEPYHEITLLEQYQCALAINDECYQVQFAOKLHKG 930
 Db 841 SKPMSRLRLAAGSAIYKLAQEPYHEITLLEQYQCALAINDECYQVQFAOKLHKG 900
 QY 931 SRLRLPEYMAICALCAKDPYKERRAHARQCLVKNINVRREYIKQHAASEKLLSLPEY 990
 Db 901 VKLLPLPEYMAICALCAKDPYKERRAHARQCLVKNINVRREYIKQHAASEKLLSLPEY 960


```

OY 991 VVPYTHLHNDPDYKODIEOLKDYKCELMFVLEIIMAKNENNSHAFIRKVENIKOT 1050
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 961 VVPYTHLHNDPDYKODIEOLKDYKCELMFVLEIIMAKNENNSHAFIRKVENIKOT 1020
OY 1051 KDAQGPDAKMEKLYTCDVAMNIIIMSKSTYVSLESPKDPVLPARFPTOPDKNSNTKN 1110
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 1021 RDAQSPDESKTNEKLYTCDVAMNIIIMSKSTYVSLESPKDPVLPARFPTOPDKNSNTKN 1080
OY 1111 YLPPEKSFPTPKRTTVNLGVANPKPLSSAGOSOTKSSRMETYS-NASSSSNPSSG 1168
    | : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 1081 YISEEIRVLLNGKPKPAGVILGVANPKPLSSAGOSOTKSSRMETYS-NASSSSNPSSG 1140
    | : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
OY 1169 RIKGR-IDSEMDHSENEED--YTMSSPLPGKSKDKRDSDLVRSLELEKRGKRKTPYTE 1225
    | : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 1141 RSRBSSSEAEETGSENEENPVRIISVTPVKNIID-----PVKNK 1179
OY 1226 EEKLGMDLTKLVQOKPKGSRKRGHTASESDE-OQPEPKRLKEDILENEDEQNSP 1284
    | : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 1180 E--INSDQAT-----QGNISDRGKKRRTVTAAGAENIIOOKTDEK-----VDESGBP 1223
OY 1285 -PKKKRGRRPKPLG-GGTPKEEPTMKTSKSKSGSPAPDEEEEROSGNT 1338
    | : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 1224 AFSKPRGRKRPKSESOGNATKNDLKNPKINKGRKR---AAGQESPGLGAGNAK 1275

```

```

RESULT 5
O8BJH6 PRELIMINARY: PRT: 824 AA.
AC O8BJH6:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE 49J10.1.2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK045159; BAC32242.1; -.
FT NON_TER
SQ SEQUENCE 824 AA; 92412 MW; 4FED723F792C71EC CRC64;

```

```

Query Match 52.6%; Score 3787; DB 11; Length 824;
Best Local Similarity 95.3%; Pred. No. 1.1e-210;
Matches 731; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

```

```

OY 625 VNSKIDGTADDEDEVPDQATRAGLELKVLSFTNHSFSAETPESLTLCKMDDEV 684
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 1 VNSKIDGTADDEDEVPDQATRAGLELKVLSFTNHSFSAETPESLTLCKMDDEV 60
OY 685 AEAALQIFKNTGSKIEEDPHIRISALLPVLHHSKSKGPPROAKVNIHCHIAFSSKEQF 744
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 61 AEAALQIFKNTGSKIEEDPHIRISALLPVLHHSKSKGPPROAKVNIHCHIAFSSKEQF 120
OY 745 AOIFEPPLKSLDPSLNLHITPLVITIGIALAPQOFAAPKSWATFIVKDLMDRLP 804
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 121 AOIFEPPLKSLDPSLNLHITPLVITIGIALAPQOFAAPKSWATFIVKDLMDRLP 180
OY 805 GKTKTKLWVPEEVSPEETMVKLOATKMWVRLGKKNHNSKSGSTLRLTLILSDGL 864
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 181 GKTKTKLWVPEEVSPEETMVKLOATKMWVRLGKKNHNSKSGSTLRLTLILSDGL 240
OY 865 TEOGKISKPDMSRLLAAGSAIVKLAQEPCHIEITLBOYOLCALINDECYQVRQVAO 924
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||

```

```

DB 241 TEOGKISKPDMSRLLAAGSAIVKLAQEPCHIEITLBOYOLCALINDECYQVRQVAO 300
OY 925 KHKGLSRLRLLEVMTATCALCAKDPVKKRRRAHQCLVKNINVRREYIKQAAASEKL 984
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 301 KHKGLSRLRLLEVMTATCALCAKDPVKKRRRAHQCLVKNINVRREYIKQAAASEKL 360
OY 985 SLLEPVVYPTTHLHNDPDYKODIEOLKDYKCELMFVLEIIMAKNENNSHAFIRK 1044
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 361 SLLEPVVYPTTHLHNDPDYKODIEOLKDYKCELMFVLEIIMAKNENNSHAFIRK 420
OY 1045 ENIKOTKDAQGPDAKMEKLYTCDVAMNIIIMSKSTYVSLESPKDPVLPARFPTOPDKN 1104
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 421 ENIKOTKDAQGPDPDKMNEKLYTCDVAMNIIIMSKSTYVSLESPKDPVLPARFPTOPDKN 480
OY 1105 FSNNTKNTLYLPPEKSFPTPKRTTVNLGVANPKPLSSAGOSOTKSSRMETYSNASSSNP 1164
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 481 FSNNTKNTLYLPPEKSFPTPKRTTVNLGVANPKPLSSAGOSOTKSSRMETYSNASSSNP 540
OY 1165 SSPGRIRGLDSEMDHSENEEDYTMSSPLPGKSKDKRDSDLVRSLELEKRGKRKTPYTE 1224
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 541 SSPGRIRGLDSEMDHSENEEDYTMSSPLPGKSKDKRDPDLVRSLELEKRPKRKAAPYD 600
OY 1225 OEKLGMDLTKLVQOKPKGSRKRGHTASESDEOQPEPKRLKEDILENEDEQNSP 1284
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 601 PEKLGMDLTKLVQOKPKGSRKRGHTASDSDEOQPEPKRHKEELLENEDEQNSP 660
OY 1285 PKKKRGRRPKPLGGTTPKEEPTMKTSKSKSGSPAPDEEEEROSGNTEDOKSK 1344
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 661 PKKKRGRRPKPLGGTTPKEEPTMKTSKSKSGSPAPDEEEEROSGNTEDOKSK 720
OY 1345 QHRVSRRAQOARAESESSAIESTOSTPOKGRGPRSKTPSPQPKNV 1391
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 721 QHRTSKRAQOARAESESSAIESTOSTPOKGRGPRSKTPSPQPKNI 767

```

```

RESULT 6
O8BJ18 PRELIMINARY: PRT: 701 AA.
AC O8BJ18:
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE 49J10.1.2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK041682; BAC31031.1; -.
FT NON_TER
SQ SEQUENCE 701 AA; 78714 MW; FAF6F0351AF7869B CRC64;

```

```

Query Match 48.0%; Score 3456; DB 11; Length 701;
Best Local Similarity 98.8%; Pred. No. 1.2e-191;
Matches 671; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 530 MYTNRNLDPGKADPFMKKFTQVLEDEKIRKOLEVLVSPSCSKQAQGCVAETTKKGN 589
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 1 MYTNRNLDPGKADPFMKKFTQVLEDEKIRKOLEVLVSPSCSKQAQGCVAETTKKGN 60
OY 590 PKOPNPPLMKKELERLAPVNIHIDTESALIKOVNSIDGTADDEDEVPDQATRAG 649
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
DB 61 PKOPNPPLMKKELERLAPVNIHIDTESALIKOVNSIDGTADDEDEVPDQATRAG 120
OY 650 LELLKVLSTFNPISFSAETPESLTLCKMDDEKAAEALQIFKNTGSKIEEDPHIRISA 709
    ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||

```

[illegible]

Db	1	LYIYASIDPAAVVALNEMKCCQMMLBSHVRELLDLHKOPTSEANCSCAMGKMLTIKKNL	60		
Qy	537	PDGKADDFMKRTQVLEDEKIRKOLEVLVSPTSCQKQAGCYREITTKRLGNPKOPTN	596		
Db	61	PDGKADQFVKFNQVLGDDEKLRSLQLELLISPTCSCKQADICVREIARKLANKOPTN	120		
Qy	597	FLEMIFLLEIAVHIDTESISLILQVYKNSIDGTADDDDEGVPDQAIKAGLELLKVL	656		
Db	121	FLEMIFLLEIAVHIDSEAIISALVKLMKNSIEGTADDEEGSPDPAIRSGLELLKVL	180		
Qy	657	SETHPISEHSAETESLCLACKMDDEKVAEAAIDIFKNTGSKIEEDPHAIRSALLPYLHH	716		
Db	181	SETHPISEHSAETESLILQCLRMEDDKVAEAAIDIFRNTGHIETDLPQIRSTLIPILHQ	240		
Qy	717	KSKGPPRQAKYALHCHTAIFSSKETOQAQIFEEPLHSLDPSNLEHITPLVITGHALL	776		
Db	241	KAKGTPPIQAAQVACHIAHIFTNKEVQADQIFEEPLRSLNADVPEQDITPLVTSIGHSML	300		
Qy	777	APDQFAAPKSMVAVTFIVKDLMDRLPGKTTTALWVDEEVSPEYVWKIQAIKMYRWL	836		
Db	301	APDQFASPKSVANFIVKDLMDLRSTGEKNGKLMSPDEEVSPEVLAHQAIKLYRWL	360		
Qy	837	LGMKNHNSGCTSTRLTLTLIHSDDLTFEODGKISKPDMSRLRLAAGSAIVKLAQEPCH	896		
Db	361	LGMKNNSKSNSTRLRLISALVSEGOITFQKRSKSDMSRLRLAAGSAITMKLAQEPCH	420		
Qy	897	EIITLEOYQALALINECYOVROVPFOKLLHKGISRLRLPLEVATICALAKQPVKERRA	956		
Db	421	EIITPEQFOCALVINECYOVARIFQFKLHKALVKLLPLLEYATIALCALAKQPVKERRA	480		
Qy	957	HAROCVKNINVRREYLKQAAVASEKLLSLPEYVBYTTHLLAHDPYVKQDIEQLKD	1016		
Db	481	HAROCCLKNISIRREYIKQNPMAITEKLLSLPEYVVPYMHLLAHDPDFRSDVDQLRD	540		
Qy	1017	VKECLMPLFELTMKKNNSHAFIRKVENIKOTKDAQGDGDKMKNEKLTVCDDVAMNII	1076		
Db	541	IKECMLPEFLEVMKKNNSHAFKMAENKLLRQDSQSPESSTNKLTYTCDCVALCVI	600		
Qy	1077	MSKSTTYSLESKPDVLPARFTTOPDKNFSNTKYNLPPEMKSEFTPEGPKPTVNLGAVNK	1136		
Db	601	NSKSAICNADSPKDPVLPMPKFTQPEKDFCNDKSYISEETRVLLLTGPKPAGVLAGAVNK	660		
Qy	1137	PLSSAGQSOYKKSRRMETVS--NASSSSNBSPPRIRGR-LDSEMHMSNED--YMS5 1191			
Db	661	PLSATGRPRPVKSTGETGSNIVNSELBPSTGRSSEQSSEAAETGVSENEENPVARIIS	720		
Qy	1192	PLPAPKSDKRDDSLVSELEKPRGAKRTVPTQEDEEKLIGMDHLTKLQDEKPKGSQSRK	1251		
Db	721	VTVPKNID-----PYKNE--INSQAT-----QGNISDQGK 752			
Qy	1252	RGHFASSEDE-QQMPPEKRLKEDLLENDEQNSP-PKKGRKGRPKPLG-GGTPKKEPTM	1308		
Db	753	RTVYAGAGENIQQATDER-----VDESGPAPAPKPRGRPRKSESOGNATKNDLNL	803		
Qy	1309	KTSKKGSKKSGPPAPEEEDDEEEROSNTE 1338			
Db	804	KPIKCKRRK---AAVGQESFGGLEAGNAK 829			
RESULT 8					
OBIXT6					
ID	OBIXT6	PRELIMINARY:	PRT: 529 AA.		
AC	OBIXT6				
DT	01-MAR-2003 (TREMBLrel. 23, Created)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DE	Similar to androgen-induced prostate proliferative shutoff associated protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				

RP SEQUENCE FROM N.A.
RC Tissue-Testis;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC039236; AAH39236.1; -
SQ SEQUENCE 529 AA; 61414 MW; 5C50973ACB920725 CRC64;

Query Match 35.0%; Score 2517; DB 4; Length 529;
Best Local Similarity 98.0%; Pred. No. 1.6e-137;
Matches 488; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAHSKTRNDGKITPPGKVEISDKISKEMVRRLKAVKTFMDMDODSEEEKELYLMA 60
DB 1 MAHSKTRNDGKITPPGKVEISDKISKEMVRRLKAVKTFMDMDODSEEEKELYLMA 60
QY 61 LHLASDFLKHGKDVRLVAVCCADIFRIYAPAPYSPDKLDMFTIRQLKGLDPT 120
DB 61 LHLASDFLKHGKDVRLVAVCCADIFRIYAPAPYSPDKLDMFTIRQLKGLDPT 120
QY 121 KSPQENRYFYLENIAMVKSYNICFELDSNEIFOLYRTLESVINNGHOKVHMVMDL 180
DB 121 KSPQENRYFYLENIAMVKSYNICFELDSNEIFOLYRTLESVINNGHOKVHMVMDL 180
QY 181 MSSICEGDYVSOELDTVLVNLVPAHKNLKOAYDLAKALKRTAQALIEYITFEFNQV 240
DB 181 MSSICEGDYVSOELDTVLVNLVPAHKNLKOAYDLAKALKRTAQALIEYITFEFNQV 240
QY 241 LMLCKTSSIDSEHVFDLILLYNIDSHLLSVLPQLEFKLSNDNEERLOVVKLLAMF 300
DB 241 LMLCKTSSIDSEHVFDLILLYNIDSHLLSVLPQLEFKLSNDNEERLOVVKLLAMF 300
QY 301 GAKOSELASONKPLMOCLYGRFNDIHVIRLECYKFAHCLMNHDLAKDLLEYLYKVRSH 360
DB 301 GAKOSELASONKPLMOCLYGRFNDIHVIRLECYKFAHCLMNHDLAKDLLEYLYKVRSH 360
QY 361 DPEAIRHDYVSIYTAAKDILVNDHLNFRVERTDCKRWRVREKAMGLAQYKKYA 420
DB 361 DPEAIRHDYVSIYTAAKDILVNDHLNFRVERTDCKRWRVREKAMGLAQYKKYA 420
QY 421 LOSAAGKDAKOIAMIKDKLHIIYONSIDRLVERIFPAQYVPHNLETTERMKCLYLL 480
DB 421 LOSAAGKDAKOIAMIKDKLHIIYONSIDRLVERIFPAQYVPHNLETTERMKCLYLL 480
QY 481 YATLDLNAVKALNEMMKC 498
DB 481 YATLDLNAVKALNEMMKC 498

RESULT 9

ID 096KV4 PRELIMINARY; PRT; 479 AA.
AC 096KV4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 49J10.1.2 (Androgen-induced prostate proliferative shutoff associated protein, isoform 2) (Fragment).
GN 49J10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z84572; CAC94787.1; -
FT NON-TER 1
SQ SEQUENCE 479 AA; 54103 MW; AB57E74AAE9656DF CRC64;
Query Match 34.6%; Score 2487; DB 4; Length 479;
Best Local Similarity 100.0%; Pred. No. 7.7e-136;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 913 DECVOYROVFAOKLHKGSLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 972
DB 1 DECVOYROVFAOKLHKGSLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 60
QY 973 LKOHAAVSEKLSLLEPEYVPTTHLHAHDPDYKVDIEOLKDVKECLMFWLEITLMAKN 1032
DB 61 LKOHAAVSEKLSLLEPEYVPTTHLHAHDPDYKVDIEOLKDVKECLMFWLEITLMAKN 120
QY 1033 ENNSHAFIRKVENIKOTKDAQOPDDAKMNEKLYTCVDAANNIMSKSTYVSLESKDPV 1092
DB 121 ENNSHAFIRKVENIKOTKDAQOPDDAKMNEKLYTCVDAANNIMSKSTYVSLESKDPV 180
QY 1093 LPARFPTQPDKNFSNTKTYLLEPEKKSFTTPCKPRTNTVNLGAVNKPPLSSAGKOSQTKSSRM 1152
DB 181 LPARFPTQPDKNFSNTKTYLLEPEKKSFTTPCKPRTNTVNLGAVNKPPLSSAGKOSQTKSSRM 240
QY 1153 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 1212
DB 241 ETVSNASSSSNPSSPGRIKGRDSEMDHSEMDYTMSSPLPGKKSDKRDSDLYRSELE 300
QY 1213 KPRGRKKTPTVTEOEKLGMDLTKLYOEOKPGSORSKRGHTJASESDQOWPEEKRLKE 1272
DB 301 KPRGRKKTPTVTEOEKLGMDLTKLYOEOKPGSORSKRGHTJASESDQOWPEEKRLKE 360
QY 1273 DILENEDQNSPPKKGKGRPPKPLGGGTTPKEEPTPMKSKGSKKSGPPAPEEERER 1332
DB 361 DILENEDQNSPPKKGKGRPPKPLGGGTTPKEEPTPMKSKGSKKSGPPAPEEERER 420
QY 1333 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQKGRGSPKTPSPQPKRV 1391
DB 421 QSGNTEOKSKSKOHVSRRAQRAESPSSAIESTOSTPQKGRGSPKTPSPQPKRV 479

RESULT 10

ID 096KV3 PRELIMINARY; PRT; 536 AA.
AC 096KV3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 49J10.1.1 (KIAA0979, isoform 1) (Fragment).
GN 49J10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z84572; CAC94788.1; -
FT NON-TER 1
SQ SEQUENCE 536 AA; 60577 MW; 3A30B33F56C81B8C CRC64;
Query Match 34.4%; Score 2476.5; DB 4; Length 536;
Best Local Similarity 99.8%; Pred. No. 3.6e-135;
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 913 DECVOYROVFAOKLHKGSLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 972
DB 1 DECVOYROVFAOKLHKGSLRLPLEYMAICALAKDPYKERRAHAROCCLVKNINVRREY 60
QY 973 LKOHAAVSEKLSLLEPEYVPTTHLHAHDPDYKVDIEOLKDVKECLMFWLEITLMAKN 1032
DB 61 LKOHAAVSEKLSLLEPEYVPTTHLHAHDPDYKVDIEOLKDVKECLMFWLEITLMAKN 120
QY 1033 ENNSHAFIRKVENIKOTKDAQOPDDAKMNEKLYTCVDAANNIMSKSTYVSLESKDPV 1092
DB 121 ENNSHAFIRKVENIKOTKDAQOPDDAKMNEKLYTCVDAANNIMSKSTYVSLESKDPV 180
QY 1093 LPARFPTQPDKNFSNTKTYLLEPEKKSFTTPCKPRTNTVNLGAVNKPPLSSAGKOSQTKSSRM 1152

```
Db 181 LPARFTQDPDNFSNNTKNLYLPPEMKSFPTPGKPTNTVLGAVNNKPLLSAGKQSQTKSSRM 240
QY 1153 ETYSNASSSSNPSPGRIKGRLDSEMDHSENEDEYTMSSPLPGKSKSDRDSDVRSLE 1212
Db 241 ETYSNASSSSNPSPGRIKGRLDSEMDHSENEDEYTMSSPLPGKSKSDRDSDVRSLE 300
QY 1213 KPRGRKTPVTEDEBEKIGMDLTKLYOEQKPKGSQSRKRGHTVASEDEQWPEEKRLKE 1272
Db 301 KPRGRKTPVTEDEBEKIGMDLTKLYOEQKPKGSQSRKRGHTVASEDEQWPEEKRLKE 360
QY 1273 DIENEDENSPPKKGRGRPRPLGGTPEKEPTMTSKGSKSGSPAPAEEREEER 1332
Db 361 DIENEDENSPPKKGRGRPRPLGGTPEKEPTMTSKGSKSGSPAPAEEREEER 420
QY 1333 QSGTEQSKSKQHRVSRRAQO-RAESPSSAISTESTPOKGRGRSKTSPSPQPKNV 1391
Db 421 QSGTEQSKSKQHRVSRRAQO-RAESPSSAISTESTPOKGRGRSKTSPSPQPKNV 480

RESULT 11
Q8N7J4 PRELIMINARY; PRT; 600 AA.
ID 08N7J4:
AC 08N7J4:
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ41012.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=uterus;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine W., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098331; BAC05286.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 600 AA; 69000 MW; 12917E7912C86BFD CRC64;

Query Match 34.4%; Score 2474; DB 4; Length 600;
Best Local Similarity 80.6%; Pred. No. 5.9e-135;
Matches 460; Conservative 64; Mismatches 47; Indels 0; Gaps 0;

QY 10 DGIITPGVKEISDKISKEVNRRLKMYKTFMDMDQDSSEKELYLNALHLASDFL 69
Db 20 DGIITPGVKEISDKISKEVNRRLKMYKTFMDMDQDSSEKELYLNALHLASDFL 79
QY 70 KHPEKDVRLVACCLADIFRIYABEAYTSPEDKLDIFMEYTRQLKLEDTKSPQFNRYE 129
Db 80 RNPKDVRLVACCLADIFRIYABEAYTSPEDKLDIFMEYTRQLKLEDTKSPQFNRYE 139
QY 130 YLEENIAWKSYNICFELEDSENEIFQLYRTLSVINNGHOKYHMHVMDLMSIIEGD 189
Db 140 YLEENIAWKSYNICFELEDSENEIFQLYRTLSVINNGHOKYHMHVMDLMSIIEGD 199
QY 190 TVSOELDLTVLNLVPAHKNLNKOAYDLAKALLKTAQALEPYTYTFEFGVLMGKTSIS 249
Db 200 GVOGELDLSTILINIPAHKNLNKOSFDLAKVLLKRYQTGTACLANFNOVLVGRSSVS 259
QY 250 DLSEHVEDLLELYNIDSHLLSVLPOLFEKLKSNDEERLQVYKLLAKMGGADESLAS 309
Db 260 DLSEHVEDLLELYNIDSHLLSVLPOLFEKLKSNDEERLQVYKLLAKMGGADESLAS 319
QY 310 QNRPVMOCYLGRFNDIVPIRLCEVRFASHCLMNHPRDLADLTLYLVRSNDPEALRHD 369
Db 320 QNRPVMOCYLGRFNDIVPIRLCEVRFASHCLMNHPRDLADLTLYLVRSNDPEALRHD 379
```

```
QY 370 VIVSIYTAACKDILLVNDHLNFEVERTLDKRRVRKEAMMGIAQIYKKTALQSAAGKDA 429
Db 380 VITITIAAKRDALVNDQLLGFVRERTLDKRRVRKEAMMGIAQIYKKTALQSAAGKDA 439
QY 430 AKQIAWTKDKLHIYTONSIDDLVRIYTAQVMPHNLLETTERMKCLYLYATLDLNAV 489
Db 440 AERYSWIKDKLHIYTONSIDDLVRIYTAQVMPHNLLETTERMKCLYLYATLDLNAV 499
QY 490 KALNEMKCONLRHGVKDDLDIKOPTASVKAIPSKVMTIRNLPDGAQDPFAKRF 549
Db 500 KALNEMKCONLRHGVKDDLDIKOPTASVKAIPSKVMTIRNLPDGAQDPFAKRF 559
QY 550 TOVLEDEKIRKQLEVLVSPSCQKQAGCV 580
Db 560 NOVLDGDEKLRQLELLISPTCSQKQADICV 590

RESULT 12
Q8VDS0 PRELIMINARY; PRT; 584 AA.
ID 08VDS0:
AC 08VDS0:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Hypothetical 65.8 kDa protein.
GN 9030416H16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA EMBL; BC021408; AAH21408.1; -.
DR MGI; MGI:1918771; 9030416H16RIK.
KW Hypothetical protein.
SQ
SEQUENCE 584 AA; 65801 MW; C7015B0681432851 CRC64;

Query Match 32.2%; Score 2317; DB 11; Length 584;
Best Local Similarity 74.3%; Pred. No. 6.8e-126;
Matches 434; Conservative 74; Mismatches 76; Indels 0; Gaps 0;

QY 525 IFSKVAVITRNLPDGAQDPFAKRFQVLEDEDEKIRKQLEVLVSPSCQKQAGCV 584
Db 1 MFGKLMITANLPDGAQDPFAKRFQVLEDEDEKIRKQLEVLVSPSCQKQAGCV 60
QY 585 KKLGNPKQPTNPLEMKIFELLERIAVPHIDTESIALIKOVNKSIDGTADDEGVPFQ 644
Db 61 KKLGNPKQPTNPLEMKIFELLERIAVPHIDTESIALIKOVNKSIDGTADDEGVPFQ 704
QY 645 AIRAGLELLVSTTHSISHSATFESLACLKMDKEKVAEALQIEFKMGKSTIEDFP 704
Db 121 AIRAGLELLVSTTHSISHSATFESLACLKMDKEKVAEALQIEFKMGKSTIEDFP 180
QY 705 HIRFALLPVLNHKKSKPPROAKVAICIAIFESSCTOPQAEPIRLKSLDSNLEHLI 764
Db 181 QIRSTLPIRLHOKAKKGTIPROAKVAICIAIFESSCTOPQAEPIRLKSLDSNLEHLI 240
QY 765 TPVLTIGHIALPADQFAAPKSWAVATFYKDLMLNDRLPQKTKTLVNPDEEVSPEYV 824
Db 241 TPVLTIGHIALPADQFAAPKSWAVATFYKDLMLNDRLPQKTKTLVNPDEEVSPEYV 300
QY 825 KIQAIKMYVRLMGKNNHSSGSTYRLTLTIHSGDITDEQKISKPMRSRLRLAAGS 884
Db 301 KQVAIKILVRLMGKNNHSSGSTYRLTLTIHSGDITDEQKISKPMRSRLRLAAGS 360
QY 885 AIVLAOEPCHHEITTEYOVLALAINDEQYOVQVFAOKLHGLSRLRLPLEYMAICA 944
Db 361 AIVLAOEPCHHEITTEYOVLALAINDEQYOVQVFAOKLHGLSRLRLPLEYMAICA 420
QY 945 LCAKDPVKERRAARQCLVKNINVRREYLRKQAAVSEKLSLLEPYVVPYTIHLAHPD 1004
```

```

|||||
421 LCAPKPKERRAHNAQCCLKNKSIRREYIKONPMATEKLSLPEVVPYMHHLAHPD 480
QY 1005 YKQVODIEQLKDYKCLFVLEIILMAKNENNSHAFIRKVENIKOTKDAQGPDKAKMNEK 1064
DB 481 FTRSDVDPOLRIKCEKLEWMELEVLMTKKNNENSHAFMKKWAENIKIKLRQASDESSTNEK 540
QY 1065 LYTQCDVAMNIIMSKSTIYSLSPKDPVLPARFFQOPKNSFT 1108
DB 541 LYTQCDVALCVINSKSAUCNADSPKDPVLPKMFQPEKVISLT 584

RESULT 13
09V646 PRELIMINARY: PRT: 1218 AA.
AC 09V646:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG17509 protein.
CN CG17509.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dushier K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garavito R.M., Geertz J., Glasner K.,
RA Glodek A., Gong F., Gorrell J.H., Gu N.S., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003823; AAF58593.1; -
DR FlyBase: FBgn003676; CG17509.
SQ SEQUENCE 1218 AA; 138887 MW; F4AGB7238A0801C2 CRC64;

```

Query Match

27.6%; Score 1982.5; DB 5; Length 1218;

```

Best Local Similarity 37.3%; Pred. No. 4e-106;
Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;
QY 13 ITPPGVKEISKDEKEMVRLLKVVKTFMDQDSEBEKELYNLALHLASDFELKHP 72
DB 4 IYPTGCRPLVEDLGTDELIRLKLTLANVLQTMDDNNLYQO-YIPALHLHLDDEFMQHP 62
QY 73 GNDVALLVACCLADIFRIYAPAPYTSQKLDQIMFMTIRQLKGLGIEDTKSPQPNKYFTLL 132
DB 63 SRDVOILLACCAVDALRYAPAPYKEDQITIKFTEFKQLHGLKDPDPSFKRYFYLL 122
QY 133 ENIAWKSYNICELEDSNEIFOTLYRFLSYVINGNHOKVAMHNVLDSSISCEGDFY 192
DB 123 ENIAVKSFMKCFELEDCOEIYODLFSITFKLVNDQHSKYVNFELVLSPIITADNIS 182
QY 193 QELDVTLVNLVPAKRNLNKQAYDLAKALKLRTQAQIIEPYITTFNOYMLGK--TJSD 250
DB 183 VELLDLILINIVEPKSNKRFACQLTEQLITFTGDALSTIKMFENRALVMDKPMTKLS- 241
QY 251 LSEHFDLILEYNDSHLLSVLPQLEFKLSNNEERLQVVKLAKMFGAKDSLSAQ 310
DB 242 ITNKIYDIYIELNRINAGILCSVLPQLEKRLSTDAERLKAATYLLSRMFSEKDSQLAKK 301
QY 311 NKPLQCYLGRFNDIHPRIECYKFAASHCLNHPDLAKDLTEYLKVSHPDEEAFIRHDV 370
DB 302 YPNLKIIFGRCDITTEPRKICVOSSAHFLNHPSLQHDITEKRLRNRHDDVEYVRHEV 361
QY 371 IVSIYTAARKDILLVND--HLNFEVERTLDRKWRVRKEMAMGLAQIYKKAQASAGKD 428
DB 362 VVAIVETAKRDEFTVLEAPDLEIIVERLDRKKYKIRDMANGLAYIKRAICEPNDLST 421
QY 429 AAK-QIAMIKOKLHIYYONSIDDLVYERIFAQVWVNHNETTERMKCLYLVATLDLN 487
DB 422 GLKVVMDIKNKLHGYKVGLEDRLVERLLITGLVPAKLPABEEMKLVLLDLADLAN 481
QY 488 AVKALNEMKCONLRHGVKDDLDLIDKPK--TDAVSKAIFSVWVITRNLPPGKAODPM 546
DB 482 ATKAVLELOKQNMKRNRYSDWIKLHSHKEFPRLVSQLSAKQANIAKLRLPRLAAEVL 541
QY 547 KKEFVLEDEDEKIRKOLELVVSPSCQAOBSQVREITRKLGPNRPPNPELEMTKFLLE 606
DB 542 TQFSNNLRKQDLICINILVKRDSRECAQDMGVLLKIGAHVQ-SNLXYNTVKMILE 600
QY 607 RIAPHIDPESIALIKOVNSIDGTADDEDEGVTDQAIRAGLELLVLSPTTHISHS 666
DB 601 RYASVWVKESITGLVLSILEQCTEKGSCEIIGISAQEGERGLKLLMLSVFSANFT 660
QY 667 AETFEISLACIKMDEKVAEALQIFKNTG--SKIEEDFPIRISALLPVLHNKSKGPP 723
DB 661 DLSLRLHLSLSYEDDYAPLVLTKLTHIGRQPLIDPPTALIDELAPVCKDFALICTP 720
QY 724 ROAKYAINCIH-----AIFSSKEIO-----FAQIFEPHLKSLDPSNLEHL 763
DB 721 KOAKHAVKCIPIFNSOSSASTGATGASASASTTQTVPIPIEIIETRLKLT-P-NCEHO 779
QY 764 ITPVITGHIHLLAPDQFAAPKSMVAFIYKDLMLNRLBQKKTKTL-----WVPDEEV 818
DB 760 RKTYTLGHIAVNMQOALFTPKNMIAIRIYAKELLIOE-VPAQRYEYELPEDSOWQAOQKL 838
QY 819 SPETVVKIQAIKMWVRLKMKNNHSGSTGTLRLITLISDGDITROGKISKPDMSRL 878
DB 839 PPDLTKIKDLAKMARWLLGLRTD-EHNAQKTFRLAFAVNGRGLDQONRLCGAKEMWL 897
QY 879 RLAAQSATVKAQOEPCYHEITTELOYOLCALAINDECYOVNOVQAKLHKGLSLRL---- 934
DB 898 RLGAACAMLKVCQEGVDOYSAEQYLOLSOLMDPVEVEIEFARKLHKGSLSLPRNC 957
QY 935 LPLEMATICALCAQDPKERRAHNAQCCLVKNINVRRETLKQHAANS-----EKLSLSP 988
DB 958 LPLDFMGLYVLAGLETERRKLLDVLNHYAETDVNKRREKLTVAAMTSPDSSTESGSLHLP 1017
QY 989 EYVVPYTIHLAHPDQYKVDIEQLKDYKCLFVLEIILMAKNENNSHAFIRKVENIK 1048

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 16:31:58 : Search time 20 Seconds

(without alignments)
3270.707 Million cell updates/sec

Title: US-09-512-581b-2

Perfect score: 7193

Sequence: 1 MAHSKRTNDKIRTPGVK.....QKGRGPRSKTPSPQPKKNV 1391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	833.5	11.6	1205	1	PD55_SCHPO
2	557	7.7	1277	1	PD55_YEAST
3	239.5	3.3	5171	1	BPEA_HUMAN
4	235.5	3.3	1875	1	MLP1_YEAST
5	229	3.2	2663	1	CENE_HUMAN
6	227	3.2	1916	1	RIF1_YEAST
7	220.5	3.1	1727	1	ALM1_SCHPO
8	217.5	3.0	1020	1	NFR_HUMAN
9	217	3.0	1790	1	USO1_YEAST
10	212.5	2.9	8545	1	ANCL1_CAEEL
11	212	2.9	831	1	NFR_RAT
12	210.5	2.9	5586	1	MDN1_HUMAN
13	209	2.9	7389	1	BPA1_MOUSE
14	208	2.9	1153	1	A3D1_HUMAN
15	206.5	2.9	699	1	NPI4_HUMAN
16	206	2.9	6632	1	UN89_CAEEL
17	205.5	2.9	2230	1	GOG4_HUMAN
18	205	2.8	2349	1	TFR_HUMAN
19	203.5	2.8	2459	1	MABP_RAT
20	202	2.8	2748	1	MDN1_YEAST
21	201	2.8	700	1	TRDN_CANFA
22	199	2.8	1359	1	ATRX_CAEEL
23	199	2.8	2464	1	MABP_MOUSE
24	197.5	2.7	3259	1	GTAN_HUMAN
25	195.5	2.7	1679	1	Y109_YEAST
26	195	2.7	488	1	CY12_BOVIN
27	194	2.7	705	1	TRDN_RABIT
28	194	2.7	1533	1	TP2A_PIG
29	194	2.7	1539	1	Y373_HUMAN
30	193.5	2.7	3210	1	CENF_HUMAN
31	192.5	2.7	2469	1	TEB0_HSVSA
32	190.5	2.6	2468	1	MABP_HUMAN
33	188.5	2.6	633	1	MLA_TETTH

34	188.5	2.6	1453	1	Y373_BOVIN
35	188	2.6	1841	1	CC12_SCHPO
36	187.5	2.6	2017	1	MYSN_DROME
37	187	2.6	1087	1	NFR_MOUSE
38	187	2.6	1311	1	APRX_DROME
39	186.5	2.6	1130	1	Y117_CAEEL
40	186	2.6	728	1	TRDN_HUMAN
41	186	2.6	2869	1	RBP1_PLAIV
42	185.5	2.6	2364	1	SPOC_HUMAN
43	184.5	2.6	3678	1	DMD_MOUSE
44	184.5	2.6	3924	1	ANK2_HUMAN
45	184	2.6	1466	1	SPA2_YEAST

ALIGNMENTS

RESULT 1
ID PD55_SCHPO STANDARD: PRT: 1205 AA.
AC Q9HFE5: O94237; 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sister chromatid cohesion protein pds5 (Precocious dissociation of DE sisters protein 5).
GN PD55 OR SPAC110.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RC STRAIN=972;
RX MEDLINE=21481773; PubMed=11598020;
RA Tanaka K., Hao Z., Kai M., Okayama H.;
RT "Establishment and maintenance of sister chromatid cohesion in fission yeast by a unique mechanism";
RL Embo J. 20:5779-5790(2001).
[2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21848401; PubMed=11859360;
RX STRAIN=972;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros K., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jags K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymoprez B.,
RA Weljens I., Vansluis E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucet M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakowski G., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
[3]
RP SEQUENCE OF 816-1205 FROM N.A.
RX Lee M., Yoo H.S., Chung K.S.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP FUNCTION.
 RX MEDLINE-21850851; PubMed-11861765;
 RA Wang S.-W., Read R.L., Norbury C.J.;
 RT "Fission yeast Pds5 is required for accurate chromosome segregation
 and for survival after DNA damage or metaphase arrest.";
 RL J. Cell Sci. 115:587-598(2002).
 CC -1- FUNCTION: Required for the establishment and maintenance of sister
 chromatid cohesion during S phase. Prevents their formation until
 esol is present. May also have a role during meiosis.
 CC -1- SUBUNIT: Interacts with esol.
 CC -1- SUBCELLULAR LOCATION: Nuclear; centromeric. Localized to chromatin
 throughout the cell cycle.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB067651; BAB71784.1; -
 DR EMBL: AL441624; CAC08560.1; -
 DR EMBL: AF049529; AAD02493.1; -
 DR GeneDB_Spombae: SPAC110.02; -
 FM Mitosis; Nuclear protein; Centromere.
 FT CONFLICT 816 818 TMC -> GTS (IN REF. 3).
 FT CONFLICT 902 908 QISLCO -> SNLIMP (IN REF. 3).
 SQ SEQUENCE 1205 AA; 138874 MW; 935DABDE0A5E30FF CR664;
 Query Match 11.68; Score 833.5; DB 1; Length 1205;
 Best Local Similarity 23.38; Pred. No. 5.2e-31;
 Matches 296; Conservative 264; Mismatches 564; Indels 149; Gaps 38;
 QY 26 ISKEWVRILKMYKTFMDNDSEEEKELYLNALSLASDFLKHGKYRLVAVACLA 85
 DB 18 LTTSSEILKRLDLGELTSLSDPT-IDRDSVLPARSLVNNMLNHLHDKGIRSTYTCIV 76
 QY 86 DIFRIYAPARYTSPDKLDMFETPRLQGLGLEDTKSPQNRVYLLNTIAWVSYNICEF 145
 DB 77 ELRLCAPDAPFTL-SQLEDIFQYILKILSLNQESTYYPQIYELLESLSNVASVLIIV 135
 QY 146 ELEDSEIFQYLTLESVYNGHNGKQVYHMHMVDLMSIIEGDTVSOELDTVNLVY- 204
 DB 136 DLPVAEELVNIFFELFEDLARKGTTKNVEFYMDIINQINLETITPAALNLIIFAOLIS 195
 QY 205 -----PAHKNLNQAYDLAKALKKPTAQATEPITTFPNOVLKMGTSISD--LS 252
 DB 196 GKGVROITGSSDSINHGPAFOLARNIFHDSADRLQRYVCQYFSDITFDSRSLDSMTTP 255
 QY 253 EHWF--DLLELYNIDSHLLSVLPQLEFKLSNDNERLOVYKLAKMGAKDSLASQ 310
 DB 256 EEFISHNLVQLMKYAPPTLLINIPFENELQAOGTQSVRLVAITVGMLL-QONALMSD 313
 QY 311 NKPLMOCYLRPNDINHPILCEYKPSHCLMNH---PDLAKDLELYKVRSHDPEAIR 367
 DB 314 YPRWSAFCGRILNDKRSVACRIKCEVANSALQNSLATSEIENVWQMSLADTDEKVR 373
 QY 368 HDVIVSVITAAKKDI-LLVNDHLLNFVREPTLDRMVRKREKMGLOIKKYVALQSAAG 426
 DB 374 VALTKTEQULTEFFKQFSVQALKMGDLRDKLAVRLQALITLSQITNRAVODLIDG 433
 QY 427 KDAAKOI-AMIKDLHIYYQNSIDRLVERIFAQYVPH-NLEETTERKCLYYIATL 484
 DB 434 VESIQWFSWLPSLLEVFVYVNDETTNAAYETICAEVLVQLYSSDQTFLNRLFLSIKYF 493
 QY 485 DLAAVKA---LDEMKGQULRHQVQ-----DLIDLKQPKTASVKAITSKAVIT 533
 DB 494 SEKAMRFILLOQRYVSELLNLYIECCRNNGVANDNEESTITNKLKVIY---DIIS 549
 QY 534 RNLPDPKADPFKKKFFQVLEDEKIRKQLEVLVSPSCSKQAGCGVREITTKKLGNGKOP 593

DB 550 SKSSNPILTATEPFFKFAE-LNDROSVMKLQ-----FFSIRSEYQVV----- 590
 QY 594 TNPLMEIKFLELIAVNIHIDTESI-----SALIKQVNSIDTADDEGEVPTDQ 644
 DB 591 -----LKSITKILFKVSVET-LSIASLECFRIFFVRSALFARNKSNVHITQLNPPVYXN 645
 QY 645 AIRAGLELLVLSFTHDISFHS-ETPESILACKMDKDEKVAEALQIFKNTGSKIEED 702
 DB 646 ELKPESEALLQHLPLIHINIVEYVIEVENIISGIESDPKVIKALSQF-----SKRKN 700
 QY 703 FPHIRSLALLVPLHNSKSGPPROKVALHCHIAFSSKEFOFOIPEPLKSLDPSLH 762
 DB 701 F-SIOTTFAELRLKLCIHGTQEQAKQATII-ATTEFERKELMITIV-----EMLEY 752
 QY 763 ---LITPLVTIGHITALAPQFAPRKSQWATFVKDILMMNDRLPCK-----KTKKLWPD 815
 DB 753 NGGLPVLMLTLGQLFTLLEE-VEKVADQTEFLVKKVI--QRFPEKYDOTHNDEKCTY 809
 QY 816 EVVSPETPVKIQAIKMYRWMLGKNHNS--KSGTSTRLTLTLHSDGLTDEGKISK 873
 DB 810 EKLDELTKCKVLAIRVLVNLRAAGTEALNIGAPIITKLKVLMLADGELSPEKPKI 869
 QY 874 DMSRLRLAAGSALYKLAQECYHEITILEQYQALAINDECYQVQVPAQKILKGLSRL 933
 DB 870 SRATRLTLASKYFLKLSIPFYAHEIDFSSVQISLCOENEDVRLFLTKLOQLK 929
 QY 934 RLPLEYVAICALCAKDPVKEBRAHARQCLVINVREYIKONAAVEKLLSLPEYVP 993
 DB 930 KLPSYPLFLFTRVADDEELKTKA-----SIWRSQVAFQKQKHDFTMEVAT 978
 QY 994 YTHLHADDDYKQVODIEOLKDYECLEFVLEITLAKNENSHAFIRKVENIKQTKDA 1053
 DB 979 YLHLHLSHNDISSIESENSIDFAYIRFYVDVYVNSENV-----PIVFHLMORIKOSYDV 1034
 QY 1054 QGPDAAKMKELIYVGVANMIIMSKS-----TYV--SLESKPDVLPAREFTOPDKN 1104
 DB 1035 --IEDG--NNITYVLSMAOKILLOVKSQNGWSLTPPKOIKLPELRLPIPSIDEKRI 1090
 QY 1105 FSNTRKYLPPEMKSFPTGPKRTTNVLGAANKPLSSAGKOSQTKSSHMETVSNASSSNP 1164
 DB 1091 FN-KIITPKMES-----QIEHALRTVPSPAKQTKTKHNAUKQKTHSSKSDK 1138
 QY 1165 SSPGRIRGRDSSMDHSENDITWSSPLPGKSKDKKDDSLVRSLEKRGKRTTYTE 1224
 DB 1139 KSSRRRK-----NEKRRLKEQNPINRVBERSSRQGIIRINYSEAPS-----SSEIRSE 1189
 QY 1225 QEERLGMDLTKL 1237
 DB 1190 EEEISEDEDEI 1202
 RESULT 2
 PDS5_YEAST STANDARD; PRT; 1277 AA.
 AC Q04264; Q04780;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Sister chromatid cohesion protein Pds5 (Precocious dissociation of
 DE sisters protein 5).
 GN PDS5 OR YMR076C OR YM9582.01C OR YM9916.15C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RX PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagsels K., Iye G., Moute S., Odell C., Pearson D., Rajadream M.A.,

RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XIII.";
 RL Nature 387:90-93(1997).
 RN
 RP FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20517447; PubMed=11062262;
 RA Hartman T., Stead K., Koshland D., Guacci V.;
 RT "Pds5p is an essential chromosomal protein required for both sister
 chromatid cohesion and condensation in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 151:613-626(2000).
 [3]
 RP FUNCTION.
 RX MEDLINE=21065936; PubMed=11137006;
 RA Panliza S., Tanaka T., Hochwagen A., Eisenhaber F., Nasmyth K.;
 RT "Pds5 cooperates with cohesin in maintaining sister chromatid
 cohesion.";
 RL Curr. Biol. 10:1557-1564(2000).
 CC -1- FUNCTION: Essential for the establishment and maintenance of
 CC sister chromatid cohesion at centromere proximal and distal
 CC regions during S phase. Also required for chromosomal
 CC condensation.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 249259; CAAB9222.1; -
 DR EMBL; 248952; CAAB8801.1; -
 DR PIR; S54451; S54451.
 DR SGD; S0004681; POS5.
 DR GO; GO:0005708; C:mitotic chromosome; IDA.
 DR GO; GO:0007076; P:mitotic chromosome condensation; IMP.
 DR GO; GO:0007064; P:mitotic sister chromatid cohesion; IMP.
 KW Mitosis; Nuclear protein.
 SO SEQUENCE 1277 AA; 147040 MW; 9DF40A5274FD9623 CRC64;

Query Match 7.7%; Score 557; DB 1; Length 1277;
 Best Local Similarity 20.3%; Pred. No. 2.7e-18;
 Matches 272; Conservative 267; Mismatches 611; Indels 190; Gaps 46;

12 KITYPGVKEISDK -ISKEEMRLKMKVKTGFMDOOSEEEKL--YLNALHLASDF 68
 8 KLEKNSPIITSDOLISTNELDLRLKALHEELASLDODNTDLTGLDKYDA--LVSRL 64
 69 LKHPGKDVRLVACCIADIFRIYAPAPYTSPPDKLKDIFMFIITROLGLEDTKSPQFNRY 128
 65 LKHKVGIRAFYACCLSDLRKLAPDAPTTA-QLTDIRKLVLSQPEGLGDOENGHYHQ 123
 129 FYLLENIAWKSYNICFELEDSNEIFQOLYR-----TLFSVINGNHQKVMHM 177
 124 TYLIRKLEYSRIVLLADPSSNNLLIELFHFYDPNKSFPARLENVIGS----- 173
 178 VDLMSIICEGTVSQOELLDTVL-----VNLVPAHKVLNQ-AYDLAKALLKRTAQAI 229
 174 --IIAEVISEFDSVPLEVLRILFNKFLTYNPNEIPEGNAVTSDCGEVSLIICDYSNNM 231
 230 EBYITTFNOVLML-----GKTSISDLSEHVFDLIELVINDISHLISLVLPQLEFKTSN 284
 232 SHHLTKYSEIHEATNDNNSRLTLVYVVKLKLVLRLMEYVPELINAIVGFIYHELSSE 291
 285 DNEERLQVVKILAKMFGA-KDSELASQNKPLMQCYLGRPNDIHPRIECCVFASHCLMN 343
 292 NELFVKRAETKLIGLITSYSDNLNFVSTHSDTFKAMISLADISPDVRAVEMTESIQIAT 351
 344 HPDLAKDLTEYIAKVAISHPEEAIIRDVYST-----VTAAKKIL--LVNDHLINVEYREPT 397
 352 REDISKEINQALAKTFIDSPRVRTSVIIFNKVPTIELMKITNKAITSILHARE-- 409

QY 398 LDKRWVRKEAMGLAQIYKKYALQ-----SAAGKDAKQJAMIKDKLLHIYONSIDRLL 454
 410 --KHREVRCLINTAKKFTYSNSLNEIERYYONKEIWEIITDIPSTLYNXYINDLINQO 467
 QY 455 VERIEAQYVPHNLLETTERMKCLLYLYATPLDIAVAKALNEMKQNLRLHQQKDLLIR 514
 468 VQSVIEFYLPEPDDKRVHRLVLSHFDDKKAFTSPFAFNARQIKISFALSKEYIDFSK 527
 QY 515 QKRTDASVAIKTSKVVV-----ITRNLPDPCAQOPMKKFTYVLEDDERIKQOLEV 565
 528 FLNNOESMSSOGPIYMNKRYNOTLQWLASGLSDSTKAIDALETIKQF--NDERIFYLNA 585
 QY 566 LVSPTSCSKQAEQVREITKCLGNPK-----QPTNPFLMIKFLLEERTAPVH 612
 586 CYTNIDIPLETFKNCYIELVSKIQTPOLFKKYNISGASIMPRD-IAKVQIILLFRASPII 644
 QY 613 IDTESIALIKOVNKSIDGTADDEGVPTDAIRAGLE--LLKVLSTHPISF-HSAET 669
 645 YVNSNISVNLNSNNS-----DAKQLDKRRILDISKVNPTLFKDQIRT 689
 QY 670 FESLIACL-----KMDDEKVAEALQIFKNIGS-KIEEDFPIIRALLPVLHNSKGP 722
 690 LKTIILKLDLDPDAEKNDNLSEELAKTKYLKASKTKLDQVDFDD--TFEFTKLYDFAVESK 747
 QY 723 PROAKYAICHIAIFESKETOFAQI--PEPLHKSIDPSNLEHLITPLTIGHIALPAD 779
 748 PETTKATKTLT-ALSKAEETLKIKIRLPDLQDKQFTSHIYIMLFFKKFPRVLND 806
 QY 780 QFAAPKSVAVATFYVKDLMLNDRLPGRKTKTL-WPDEEVSPETVW---KIQAIKMWVR 834
 807 D-----STDIIISYLIEVLLSNQVGDSSKEIDWDSLSIDPTKYSAINKVFYTLKLPFN 861
 QY 835 WLLGM-----KNNHSGST-SFLRLTLTLLHSDGL-TPQGIKSRDWS-----RLRLAG 883
 862 KLRSTAPDVPRDELSEFTEKTKMLFEVYLASGGEISFENKEFYTPSNYOTKRLCVAG 921
 QY 884 SAIVKLAOPRCYHEIITTEPOYLALAINDECYQROYAPQAKLHGKLSRLRLPLEYMAIC 943
 922 IGVLTAKRISNLNMFIRPSDIIKLNIVDESLPRPKRTFLBQDKQVYANELISIRFLPV 981
 QY 944 ALCANDPVKERRAHARQCLVKNINVRREYLKQHAANVSEKLSLLEPVVYPTIHLANDP 1003
 982 PFTAAEPDVELTTTKIWI--NFTGLMSFKR-----GTIFERALPRLIHAIHNP 1030
 QY 1004 DVYKVDIE-----QLADVKECILMFVLEIIMAKNENSHAIIRKAVENIK-----QTK 1051
 1031 DIVGGIDSEGDAYLNAALTTATIDYLLFEYPSIAAQENFSLLYLSERKVNQOKLVEDET 1090
 QY 1052 DAQGPDDAK-----MNEKLYTCDVAMNIIMS-----KSTYVS-----LESKDPVLPA 1095
 1091 EEEGQKEAPKPKKHRYGQKMTIIGELSQMILLNLKEKKNMQHSATPGKLNLPDLFKP- 1149
 QY 1096 RFTTQPDKNFSMTKNYLLPREMKSFPTPKPKTKTNVLGAVNKPLSSAGKOSQTKSRMETV 1155
 1150 -FATVQEAQLS-FKTYIPESLT-----EKIQNNIKAKIGRLIHSQTRQRLQKRLLAH 1201
 QY 1156 SNASSSSNPSSPGRIKGRIDSSEMDHSEMEDYTTMSSPLPGKSD-KRDDSILVRSELEK 1214
 1202 ENNESQKKKKKHHARSQADDEGDGDRSDSDSDSYSPSNKNETKKGHENIV---MKKL 1258
 QY 1215 RGRKTKPVYEOEEKLGMDL 1234
 1259 RVRKEVDYKDDSD---DDI 1274

RESULT 3
 BPEA_HUMAN STANDARD: PRT: 5171 AA.
 AC 094833; Q8N1T8; Q8N8J3; Q8WKK9; Q96AK9; Q96Q05; Q9H555;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bulbous pemphigoid antigen 1, isoforms 6/9/10 (Trabeculin-beta)

DE (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)
 DE (Dystonia musculorum protein).
 GN BPAG1 OR DMH OR DT OR KIAA0728.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=21839111; PubMed=11751855;
 RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
 RT "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
 RT including the domain structure closely related to MACF (microtubule
 RT actin cross-linking factor).";
 RL J. Biol. Chem. 277:6682-6687(2002).
 RN [2]
 RP SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).
 RC TISSUE=Dundenum;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Melek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fehy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain, Placenta, and Tongue;
 RA Niomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Isogai T., Nishi T., Ota T.,
 RA Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Kikuchi H., Masuno Y.,
 RA Nagai K.;
 RT "NEDO human cDNA sequencing project.";
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain;
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [5]
 RP SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:277-286(1998).
 RN [6]
 RP SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).

RA Smith M.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP DISEASE.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=92011493; PubMed=1717441;
 RA Sawamura D., Li K., Chu M.-L., Uitto J.;
 RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
 RT deduced from cloned cDNAs predict biologically important peptide
 RT segments and protein domains.";
 RL J. Biol. Chem. 266:17784-17790(1991).
 CC -1- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
 CC intermediate filaments to the inner plaque of hemidesmosomes. The
 CC proteins may self-aggregate to form filaments or a two-dimensional
 CC mesh (By similarity).
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named Isoforms=10;
 CC Name=6; Synonyms=EA;
 CC IsoId=Q94833-2; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q03001-1; Sequence=External;
 CC Name=2;
 CC IsoId=Q03001-2; Sequence=External;
 CC Name=3; Synonyms=1e;
 CC IsoId=Q03001-3; Sequence=External;
 CC Name=4;
 CC IsoId=Q03001-4; Sequence=External;
 CC Name=5;
 CC IsoId=Q03001-5; Sequence=External;
 CC Name=7; Synonyms=EB;
 CC IsoId=Q8WXK8-2; Sequence=External;
 CC Name=8;
 CC IsoId=Q03001-6; Sequence=External;
 CC Name=9;
 CC IsoId=Q94833-3; Sequence=VSP_005068, VSP_005069;
 CC Name=10;
 CC IsoId=Q94833-1; Sequence=VSP_005066, VSP_005067;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
 CC cultured keratinocytes.
 CC -1- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
 CC [MIM:600088], an autoimmune subepithelial skin blistering disease.
 CC -1- SIMILARITY: Belongs to the plakins or cytolinker family.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 29 spectrin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF400226; AAL62061.1; -;
 DR EMBL: BC016991; AAH16991.1; -;
 DR EMBL: AK055189; BAB70870.1; -;
 DR EMBL: AK094883; BAC04449.1; ALT_INIT.
 DR EMBL: AK096713; BAC04448.1; ALT_INIT.
 DR EMBL: AB018271; BAA34448.2; -;
 DR EMBL: AL137008; CAC12899.1; -;
 DR HSSP: P02631; IRR0.
 DR Gene: HGNC:1090; BPAG1.
 DR MIM: 113810; -;
 DR MIM: 600088; -;
 DR GO: GO:0005200; F:structural constituent of cytoskeleton; ISS.
 DR GO: GO:0045104; P:intermediate filament cytoskeleton organiza. .; ISS.
 DR Interpro: IPR002048; EF-hand.
 DR Interpro: IPR003108; GAS2.
 DR Interpro: IPR002017; Spectrin.

DR Pfam: PF00036; eHand; 2.
DR Pfam: PF02187; GAS2; 1.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00435; spectrin; 28.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; Eph; 2.
DR SMART: SM00243; GAS2; 1.
DR PROSITE: PS00018; EF_HAND; 2.
KW Antigen: Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW Calcium-binding; Alternative splicing.
FT REPEAT 264 341 SPECTRIN 1.
FT REPEAT 349 444 SPECTRIN 2.
FT REPEAT 450 550 SPECTRIN 3.
FT DOMAIN 564 616 SH3.
FT REPEAT 935 1031 SPECTRIN 4.
FT REPEAT 1481 1581 SPECTRIN 5.
FT REPEAT 1715 1814 SPECTRIN 6.
FT REPEAT 1930 2008 SPECTRIN 7.
FT REPEAT 2071 2174 SPECTRIN 8.
FT REPEAT 2184 2282 SPECTRIN 9.
FT REPEAT 2294 2382 SPECTRIN 10.
FT REPEAT 2404 2502 SPECTRIN 11.
FT REPEAT 2513 2611 SPECTRIN 12.
FT REPEAT 2839 2940 SPECTRIN 13.
FT REPEAT 2950 3047 SPECTRIN 14.
FT REPEAT 3059 3156 SPECTRIN 15.
FT REPEAT 3277 3376 SPECTRIN 16.
FT REPEAT 3386 3484 SPECTRIN 17.
FT REPEAT 3495 3594 SPECTRIN 18.
FT REPEAT 3716 3816 SPECTRIN 19.
FT REPEAT 3825 3925 SPECTRIN 20.
FT REPEAT 3935 4035 SPECTRIN 21.
FT REPEAT 4044 4142 SPECTRIN 22.
FT REPEAT 4153 4250 SPECTRIN 23.
FT REPEAT 4262 4362 SPECTRIN 24.
FT REPEAT 4372 4470 SPECTRIN 25.
FT REPEAT 4480 4578 SPECTRIN 26.
FT REPEAT 4596 4684 SPECTRIN 27.
FT REPEAT 4732 4790 SPECTRIN 28.
FT REPEAT 4798 4810 SPECTRIN 29.
FT CA_BIND 4834 4846 EF_HAND 1 (POTENTIAL).
FT DOMAIN 126 160 EF_HAND 2 (POTENTIAL).
FT DOMAIN 312 377 COILED COIL (POTENTIAL).
FT DOMAIN 402 447 COILED COIL (POTENTIAL).
FT DOMAIN 511 546 COILED COIL (POTENTIAL).
FT DOMAIN 706 813 COILED COIL (POTENTIAL).
Query Match 3.3%; Score 239.5; DB 1; Length 5171;
Best Local Similarity 18.5%; Pred. No. 0.006; Mismatches 448; Indels 467; Gaps 66;
Matches 263; Conservative 241; Mismatches 448; Indels 467; Gaps 66;
QY 22 ISDKTSKEWVRLKMYVTFMDMDQSEEEKLYLNLHLASDFELKHPGKDYRLAVA 81
DB 2395 VKEEVLKAEKLSRQGLGILSKFEVDQKAE-----NHVQHLSQACASHQ----- 2438
QY 82 CCLADIPIRIYAPADPYTSPDKLIDFIMFTTRQLKGLDETKSPQENRYFYLLENIAMVKS 141
DB 2439 -----FOOMSDFQAMLDTRKKEQN-----KSH 2461
QY 142 NICEPELEDSNEIFQOLYRTLPFSVINNGH-----QKVMHNVADLMSSITICGDRV----- 191
DB 2462 PLSAKLD-----VLESLLDKHDKFSTLTAAOSHMY-----EKTIAEGENLLKT 2505
QY 192 -----SOELDIYLVNLYVPAHKNL-----NKQAVDLAKAL-LKRTAQALIEPYITTFNQ 239
DB 2506 GSEKAKALQALQNTIKTMDPTFNKQVKEKEREKLSLEKALKYKQOVETLMPWIDKCOQN 2565
QY 240 YLML-----GKTSIS-----DLSEHVFDLILELYNIDSHLLSVLPOLERKLSN 284
DB 2566 LEEIKFCLDPAGEISIAKLKSLQKEMDQH-FGMV-ELIINTANSILSYC-EIDRGVYTD 2622

QY 285 DNEERLYOVKLLAKMGARD-----SELASQNKPLMOCYGLRGNDIHVP-- 329
DB 2623 ENKSLIQKVDVTEQHSKSKFCLENNMTQKFEHQEVSKEKSNOLQCAKQQL-DIHSLGS 2681
QY 330 -----RLECVFASHCILNHPRLADLTLEYLKVSRHDEBEAIRHYVYSITYAAK 380
DB 2682 QAVSNKYTLMLQTOOKSLQALKHQYDLAKRLADLVEASDSKGT--SDVLLQVETIAOE 2739
QY 381 DLLVNDHLNVRENTLTKRMVRKREAMVGLQIYKKAVALSAAGKAADKAOLAMIKDL 440
DB 2740 -----HSTLSQVDEKCSFLEKL 2758
QY 441 LHI-YYONSIDRLIVERIFQAQWVPHNETTERMKCLYLYATLD--LNAKVALNEMAK 497
DB 2759 QGIGHFRNTI-----REMFQ-----FAEPDELDMSAPV----- 2789
QY 498 CONLNRHOVKLLDLIKQPKTDASVKAIFSKVAVITRNLPDGKADQEFKKFTQVLEDE 557
DB 2790 -----RDAGTLQKQKET-----IKAFLEKLEALMASNDANKTKCMILATEESPDLY 2837
QY 558 KIRKOLEVLYSPFCSC-----KQAGCYREITKKLGNPQPTNPFLEMIFLELRI 608
DB 2838 GIKRDEAL-SKQCNKLLDRAQAREQVGTIKRLEEFYSKLE-----PSLLQRA 2888
QY 609 -----AFVHIDTESIALI-----KQVKSIDGTADDED---EGV-----PTDQ 644
DB 2889 EEHEESQGVGKMETETINQOLMKFKVQKEBELIEPLQKQDVNWLQGLIOGSAKSTSQ 2948
QY 645 AIRAGLEILKV--LSFTNPISFHSATFESSLACLKMDKQVAAALQIFKNTGKIEED 702
DB 2949 GLEHDDVDVNAEMKTLNKKVAQRAAOLQALHLHGCFRD--ALESLISMVAVTEELVANO 3006
QY 703 PPHISALLPVLIHNSKKRPPROAKYAINCHAIIPSKEP-----QFOIIEPLMK--- 753
DB 3007 KP-----PSAEFKVKAQIQEQKILQRLDLDRKSTVEVIRREGKIAITPAEPADKVI 3059
QY 754 -----SLDPSNLHLITPLVT-----IGHILAPDQF---AAPKSVATFIYKDLMDNR 802
DB 3060 LKQSLSDSRWELALNKATRRNROLEGISVA-QQFHELEPLNEMLT--IEKRLVNE 3116
QY 803 LPQKTKTKLMPDDEVSPETWVKIQAIKMMVWMLQMKNNHSK-----SGTSTRLTLTT 856
DB 3117 PIGTQASKL---EQQIAQKALEDDII-----NHKKHLHQAVSIGQSILVSS 3161
QY 857 ILHSGDLTEQCKISKPMDSR-----LRLAASAIYKLAQEPCEYHETITLQYQCL--AL 909
DB 3162 --REDKDWQ-----SKLDFSQWYVTEIQKSHSRSELQQAQCNKIFGEDEVELNMVLN 3215
QY 910 AINDSC--YQVRQVFAQKHLKGLSLRLPLEYMAICALCAKDPVKERRAHAROCVKNIN 967
DB 3216 EYHDKLSKLSVDYSTEGMLKQOSELRV-----LQEDILRKQNVQDALLNGL- 3263
QY 968 VREERYLKQHA-----VSEKLISLPEVYVPTYIHLAHDPYVAVQVODIQL-DVKECL 1021
DB 3264 ---ELLKQTTQGEVILIQKLEAIKARY-----KDTIKLSTQVAKTL 3302
QY 1022 WFLVELILAKNNNSHAFIRKRVENIKQTKDAQGDDAKAMNKLYTVCDAVANIIMSST 1081
DB 3303 EQALO--LARRLHSTH-----BELCTWLDKVEVELISYET 3335
QY 1082 TVSLSPKQPVLPARFPTQPDKNFSNTKNYLPPKMKSFPTPKPPTVNLGAVNPLLSA 1141
DB 3336 -----QVLKGEASQAQMR-----PKELK----- 3354
QY 1142 GKQSQTSSRMETVSNASSSSNPSPGRIGRLKGLDSSEMDHSENEDYTMSSPLPGKSKDR 1201
DB 3355 -KEARKNNKALLDSLNEVSALLLELVPMWRAREGLE--KWAEDENERIYVSDTITQKVEEV 3411
QY 1202 DSDSLVRSLEKPRGRKTPYVTEQBEKL-GMDDLTKIVQEQPKRSQ-----NSR 1250
DB 3412 DAAILRSQQFDQADAELSMWITETEKRLMSIADI-RLQDDQSAOLQOVQKTFYTWELRHR 3470
QY 1251 -----KRGH---TASESDQQWPEKRLKEDILNED 1279


```

Db 1343 -----SKLSQSLREQVNSLRDA-----KNVLNSLSEANNAIEELQNAKVAQ 1385
QY 1127 TTNVCAVANKPLSSAGKQSTKSMRETVSNASSSSNPSPGRIGRLDSSMDHSENED 1186
Db 1386 GNNQLEAIRKTLQEDAEKASRELQAKLE-----ESTSYESTINCL-----NBE 1428
QY 1187 YTMSSPLPGKSKDKRDDSLVRSLELEPRG-RKKTPVYEQEELKMDLTIKLYQKPRG 1245
Db 1429 IT-----TLKELEKORQIQOQLQATVANSQO-----NDLSNIVESENK--- 1465
QY 1246 SQRSRKRGHTASESDQOWPEEK--RLKEDILNEDQNSPP-----KKGKRGPRPKP 1296
Db 1466 -----KSFEEKIKFKIKETQGVNKKILAEQERLNQPSNIMEEKKKKWESEHDE 1516
QY 1297 LGGGPKKEEPTMK-----TSKKSCK---KKSPPAPEEEEEEER-----OSG----- 1335
Db 1517 VSQKIREAEALKKRIRLPTTEKIKNKIIEKKKEELEKEEKEEVEERIKSMQSGEIDVYL 1576
QY 1336 -----NTEOKSKSKOHVRSRAQOARS-PSSASIESNQ 1368
Db 1577 RKOLEAKVOEKOKELENEYNNKLQELKDVPHSHSIDSE 1616

RESULT 5
CENE_HUMAN STANDARD: PRT: 2663 AA.
ID CENE_HUMAN
AC 002224:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RX Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RX "CENP-E is a putative kinetochore motor that accumulates just before
RX mitosis.";
RT Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RX Thowar D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RX "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RX microtubule motor.";
RT EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RX Chan G.K.F., Schaar B.T., Yen T.J.;
RX "Characterization of the kinetochore binding domain of CENP-E reveals
RX interactions with the kinetochore proteins CENP-F and hBUBR1.";
RT J. Cell Biol. 143:49-63(1998).
RN [4]
RP FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
RP KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
RP OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
RP AND/OR SPINDLE ELONGATION.
RP -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
RP CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
RP QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
RP -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 215005; CAA78727.1; -.
DR PIR: S28261; S28261.
DR HSSP: P17119; 3KAR.
DR Genbank: HGNC:1856; CENPE.
DR GK: Q02224; -.
DR MIM: 117143; -.
DR GO: GO:0005699; C.kinetochore; TAS.
DR GO: GO:0005634; C.nucleus; TAS.
DR GO: GO:0008350; F.kinetochore motor activity; TAS.
DR GO: GO:0008067; P.DNA replication and chromosome cycle; TAS.
DR GO: GO:0007079; P.mitotic chromosome movement; TAS.
DR GO: GO:0007080; P.mitotic metaphase plate congression; TAS.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS00467; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEC138B0C8CB8 CRC64;

Query Match 3.2%; Score 229; DB 1; Length 2663;
Best Local Similarity 18.2%; Pred. No. 0.0077;
Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;

QY 17 PGVKEISDKISKENEVRRIKIVYKTFMDQDSE-----EKELYNLALHLASD 66
Db 333 PYVNEVS--DDEALLKRYR--KEIMDLKQLEVSLETRAQAMEKD--OLAOLLEBK 383
QY 67 FFLKHPGKGVRLVAVACLAIDFRIYAPAPYTPSPKLDIF-MFTIRLQKL-EDTKSPQ 124
Db 384 DLQK-----VQNKETNLDTMLTSSSTLTQOELKRR 417
QY 125 FNRFYLLLENIAVWKSYNICEFELDSNEIFQLYTFLSVINGNHQVHHMVDLMSSI 184
Db 418 KRRVWCICLKIKKMKNSN-----YADQNIPIPN-ITTHKLSINLKEI 461
QY 185 ---ICEGPTVSOELDVLV-VNLVPAHKNLKQAVDLAKALKIKRTAQAIPEYITTFNOV 240
Db 462 DESVCSDESDFVSNTLDTSEIEMNPATKILNDENIE----- 497
QY 241 LMLGKTSISDSEHFIDLILELYNIDSHLLSVLPQLEFKLSNDNEERLQVYKLLAKMF 300
Db 498 -----SEINSLRADYDNLVL--DYQLRTREKEEMELKLEKNDLDEFLEKRT-- 545
QY 301 GAKDSELASQMKPLMQCYLGRFNIDHVPRIELCVAFASHCLMNHDPDLAKDLTEYLKVRSH 360
Db 546 --KQDEQ-----LIHEISNLKNLYKIHREYVNO 571
QY 361 DPEAIRHDVIVSYTAAKKDILAVNHLNFVRERTLDKRRKREKAMGLAQYKRYA 420
Db 572 DLENEL-----SSKVELLREKEQIKKLEVIDSOR--LENIKMPLS-----YS 613
QY 421 LOSAAGKDAKQAIAMIKKLIHYYQNSIDRLLVERIFAQYMPVPHNLETTERRMKCLYLL 480
Db 614 LESTEDPKQMKOTLPDAFTV-----ALDAKRESAFLNSENLEKMKELATTT 661
QY 481 YATLDLNAVKALENEMKCONLIRHQVKDLDLIRKPKDASVKAIFSKVAVYTR----- 534
Db 662 YKQNE-----NDIQLVQSQLEAKKKMQVDLEK-----LQSAFNRITRLTSLDSKV 708
QY 535 -----NLPDGKADPFKKTFTQVLEDEKTRKOLEVAVSPSCQAQEGCVREITKLG 588
Db 709 PKDLICNLLEGKITDLOELKELNKEVEENALREEV-IILS--ELKSLPSEVERLERKEI- 763

```

```

QY 589 NPKOPINPELEMIKFLERLAPVHIDTES-1SALIKOVNKSIDGTADDEGEVPTDOAI- 646
DQ 764 ---ODKSEELHITSEKDKLFSEVHVESHVQGLLEBKIKTKDDLATQSNKSTDOEFQ 820
QY 647 -----RAGLELKVLSTHPIFSATFESFLACLMD----- 680
DQ 821 NFKTLHDFEOKYKMWLEENERNNOETVNLK-----EAOQFSDSLALTELSTYKT 872
QY 681 ---DEKVAEALQIFKNTGSKIEDEFPHRSALLPVLHKS-----K 719
DQ 873 QELQEKTRVQERL--NEMQOLKEQLENRSPLOTYERETLTLEKLOQLEEVKTLTQE 930
QY 720 KSPRQAKVAIHCHIAIFSSKEFOAOFEPRLKSLDPN-----LEHLTPVLTGHIA 774
DQ 931 KDLKLOESLQI-----ERDOLKSDIHDTVMNIDTQOLRNALLESILKHOEFTINTL- 983
QY 775 LLAPOGAAPWKSQVATFIYKDLMDRLPGKTKTKMWDEVEVSPETWIKQAIKMWV 834
DQ 984 -----KRISEVSRNLHM-----EENTGETKDEFQ-----Q 1010
QY 835 WLLGMKNHSGSTSTLRLTLTLHSDGDLTEOGKISK--PDMRSLAAGSAIVKLAQ- 891
DQ 1011 KMWGIDKODLEAKNT-QTLTADYKONEIEQOKRIFSLQEKNELOQMLESVIAEKEQL 1069
QY 892 -----EPCHYEITTEOYOLCALAINDCYOVROVPAOKLHKLRLPLEYMAICALC 946
DQ 1070 KTDLEKENIEMTENOELRL-----LGDLEKKOEIYAOEKRNHAIKK---EGELSRCTDRL 1122
QY 947 A--KDPKERRAHARQCLVNIWVRREYLKQNAVSP-----KLISLLEPVVYTI 996
DQ 1123 AAEVEKIKKESQOLOEQOOLVNOEWSMOKKINTEINIKMELKKELTLEMETERL 1182
QY 997 HLIAH--DPDYVYQDI--EQLDVKECIMEVLEILMAKENNSHAFIKKAVENIKOTDAQ 1054
DQ 1183 ELAQKLENEVEVKSTIKERKVLKE-----LQKSFERHDLNGYIRFEIATGQRTKEEL 1237
QY 1055 GPDDAKNEKLYVCDVANNIIMSKSTYTSLESPKDPVLPARFPTOPDKNFSNTKNYLP 1114
DQ 1238 KIAHILHKEHOETIDELIRSV--SEKT-----AQIINTQDLEKSHRK--LQE 1280
QY 1115 EMKSFTPGKPTNTVNGAVNKPSSAGKOSQTKS--SRMEYVSNMSSSNPSPGRKIG 1172
DQ 1281 EIP-----VLHEQELLPYKVKVSETOETMNELELLELLESTTQDSTTLARI-- 1326
QY 1173 RLDSSEMDHSE--NEDYTMSPLPGKSKDKRDDSDVLESELEKPRGRKKTVTYE----- 1224
DQ 1327 -----EMERLRLMEKQESQEBELKSLKERDNLKTIKALEVKNHDOLEKHIIRETLAKIOE 1381
QY 1225 -----QEBKLGK-----DDLTVLVOBQ--KPK-----GSQSRKRGH-----TA 1256
DQ 1382 SOSKOEQSLMKERKDNETTVISEMEOFKPKDSALLRIETEMLGLSKRLQESHDENKSA 1441
QY 1257 SESDEQO-----WPEEKRLKEDI-----LENEEDQNSPPKGRGRPP--KPLGGCTP 1302
DQ 1442 KKKDDDLQRLQEVLOSQESDQLEKNIKELVAKHLETEELKVAHCCLEKQOETINELRVNLS 1501
QY 1303 KEEPTMKTSRK-----GSKRKSQPPAPEEEDER--EROSGNETOKSKSKOHVRSRAQQ 1354
DQ 1502 EKEFEISTIQKOLEAINDKQNKIOETIYEKEBQNLKQISEVOENVNELQKPFERRAKXD 1561
QY 1355 RA-ESPESSAIEST 1367
DQ 1562 SALQSIESTKMLELT 1575

```

RESULT 6

```

RIF1_YEAST STANDARD; PRT; 1916 AA.
AC P29539;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RIF1 protein (Rap1-interacting factor 1).

```

```

GN RIF1 OR YBR275C OR YBR1743.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB320;
RX MEDLINE=92249772; PubMed=1577274;
RA Hardy C.F.J., Sussel L., Shore D.;
RT "A Rap1-interacting protein involved in transcriptional silencing and
telomere length regulation."
RL Genes Dev. 6:801-814(1992).
RN [2]
RP SEQUENCE OF 1-1096 FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=94378722; PubMed=8091861;
RA Holmstrom K., Brandt T., Kalliesoe T.;
RT "The sequence of a 32,420 bp segment located on the right arm of
chromosome II from Saccharomyces cerevisiae."
RL Yeast 10:S47-S62(1994).
RN [3]
RP SEQUENCE OF 1005-1916 FROM N.A.
RC STRAIN=5288C;
RA Aigle M., Bacle M.C., Barthe C., Bileau N., Crouzet M., Dignon F.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP FUNCTION. AND SUBUNIT.
RX MEDLINE=97242430; PubMed=9087429;
RA Wotton D., Shore D.;
RT "A novel Rap1-interacting factor, Rifp, cooperates with Rifp to
regulate telomere length in Saccharomyces cerevisiae."
RL Genes Dev. 11:748-760(1997).
CC -!- FUNCTION: Involved in transcriptional silencing and telomere
length regulation. Its role in telomere length regulation results
from either a block in elongation or promoting degradation of the
telomere ends. Loss of Rif1 function results in derepression of an
HMR silencer, whose ARS consensus element has been deleted, and in
the elongation of telomeres. Rap1 may target the binding of Rif1
to silencers and telomeres.
CC -!- SUBUNIT: Interacts with Rif2 and Rap1 C-terminus.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; Z36144; CAAB5238.1; -
DR EMBL; X76053; CAAS3638.1; -
DR EMBL; X66501; CA47121.1; -
DR PIR; S46157; S46157.
DR SGD; S0000479; RIF1.
DR GO; GO:0005696; C:telomere; IDA.
DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; IMP.
KW Telomere.
FT CONFLICT 580 580 C -> S (IN REF. 1).
FT CONFLICT 732 732 T -> A (IN REF. 1).
SQ SEQUENCE 1916 AA; 217959 MW; 17B5487151CFE665 CRC64;

```

Query Match 3.2%; Score 227; DB 1; Length 1916;

Best Local Similarity 17.4%; Pred. No. 0.0062;

Matches 316; Conservative 267; Mismatches 612; Indels 624; Gaps 80;

```

QY 3 HSKRTNDGKITV-----PPQVKEISDKI-----SKENVRRLKAV--- 38
DQ 183 YKFRVNPRLALEVWGEIHGLVDNBSVSFEKELIGGLGLQESBDVYARREFVATP 242
QY 39 -----VKTFDMDQDSEEEKELVNLNA-----LHLASDFFLKHPGQDVLLVACCL 84

```



```

Db 243 NNIIPLTNTKNNVEYDOKENILIVNIESIIEICIPHLOIAODPTLLSSSEKKNPVYI----- 298
QY 85 ADIFRITVPEAPYTS-----PD---KLDIF-----MFTR 112
Db 299 ---RLVYOIVRFFSAINSPKRIYKWLTRPDLVKLVYIYRMWTGALRNENSKNIIITA 354
QY 113 QLKGEJEDRSQOFNFFYLENI-AMVKSYNICFELEDSENEFFQOLYRTLFESVINGNHQ 171
Db 355 QVSEFRDEK---FGFELISNEIKPIISTEIMEINSHNLIYEKL-----LIRGLIS 405
QY 172 RVHMHVMDJMS-----ICBGDTVSOELDTVLVNLVPAKNNKQAYDLAKAL 222
Db 406 KYPKMIETVTSWMLGEVLPRIIIDEIYSMKILITSIVLL-----ELL 450
QY 223 KRTAOAIEPIYITTFNOVLAMGKTSISDISEHVPULIELY---NID---SHLLSVIP 275
Db 451 KKCIDEVDEHERIY--QCIMLSPVCETIPEKFLSKPLNSYDSANILDKVYTIHLLTQOIK 508
QY 276 OLEFKLGSNDNEERLOV-VKLLAKMF--GAKDSELASQNKPLM-----QCYLGRPNDIHV 327
Db 509 ---NYIVAKNDKIAMDLMLSMGLIYDSGKRYDILTESKNKWPDLNLCFTNN---HP 562
QY 328 PIRLECVR--FASHCL-----MNHPLAKDLTEXIK-----VRSHEEALRHVY 370
Db 563 KTRLMISIKVMRIITCYICTKISQKQEGNKSLLSLRTPQWTLRYVNDPSAREGIIYHL 622
QY 371 IVSIVTA--AKKDIL-----LVNDHLNLFVRERTIDKRRVR-----KEA 408
Db 623 LGVYVTAFTSNKNLSTDMFELEMDLITPIYEDYVFKDSIHQVLFVTLHLIGKNA 682
QY 409 MMGLAQIYKKY-----ALQSAQAQDAK---QJAMWK---DKLHIYQNSIDRLVY 455
Db 683 DYALEKRTKHHHPMSVIASBEGVAKLIDISSLPQI--IKREYDKMKVYFQI-----V 733
QY 456 ERIEAOYVPHNLETTERMKCLYLLATFLDNAVALNEMKQCNLLRHQVKOLLDLIKO 515
Db 734 EVAISNVMLADHLILTS-----LKHLPD----- 757
QY 516 PRTDSVKAIFSKVAVITRNLPDPGKADPFMKFTQVL-----EDEKIRKOLE 564
Db 758 RKOQTHLESFSSILIKVQNNKDPPIFRDFGAVSSFEVYTFDLFLRKNDSLVNPNQO 817
QY 555 VLVSPFCSQKQAGCVREITKKIGNPQPTNPFLMKIKFL-----ERIA 609
Db 818 ISKVGISOGNMTLDLKVIRKARN---ETSEFLIEKLELDKKEVYAQNWGSTLL 874
QY 610 PVHIDTESTALIKOVNKSIDGTADDEDECVPTQAIRAGLELKVLSFTPHISFSAET 669
Db 875 PNINISREFEOSIANIYNK-----VENENSIENFLDLCLKLSF--PVNL----- 915
QY 670 FESLLACLKMDKEKAAEALQIFKMTGSKIEEDF-----PHIRSALLPVLHHS 718
Db 916 FTLLHVSMNSNNNFYFTQSYVSKNE-NKLVNDLITLKTSLPGNPELSGGLPFL----- 970
QY 719 KKGPRQAKY--AIHCIIA-----IFSSKE- 741
Db 971 ---RRMKFMDILEYCIHNSNPNLNSIPDLNSDLLKLPSSRSASYFANIKFKCSQO 1025
QY 742 -----TOFAOIFEPRLKSLDPSNIE-HLITPLVTICHIALLPADOFA 782
Db 1026 ILLVWMLLKGOOLEONFSEIENVLQNASD-SELEKSEIIRRELL--HLAMANDIE-- 1079
QY 783 APWKSVMVATFIYKDLMLN--DRLPKKTTKLWVPDEEVPENMVITQAIKMMVNRMLGAK 840
Db 1080 -PLFSGLLNFCIKNNMAHDLDFCGNMVTSVLP--KISPELLKILTYK----- 1125
QY 841 NNHSKSGSTLRLTLTIHSDGD---LTEQGISKSPDMSRLIAGAIVKIAOEPCHY 896
Db 1126 ---ERKNGLKLAIVIEKIENGDDDIYELLEKIITIIQKE-----IOLEKLEP--- 1170
QY 897 EITTEYOQALAIINDECYO-----VROVFAQKLHKHGISRLPLEYMAICALCAKD- 950
Db 1171 ---LLVFFLNFVSSNMQKHKKSTWMLRELVLVLTFLPLRSAAKKFFSMLISILPPNEN 1226

```

```

QY 951 -----VERARAHQCLVKNINVRRELYKO-----HAASEKLSTL 986
Db 1227 YQIDMVNLLIDLKSHNKKRFKDRTY--NATLKTIG---KWIQESGVHOGDSSEKIEA 1281
QY 987 LPE---VY-----VPYTIHLAHPDVYKODIEOLKDVEC 1020
Db 1282 IPDTSKMIIPCEGSEKNLSNLOARKVDSODIQVPAT-QGKKEPSSIQSSQISAND-SDS 1339
QY 1021 LMFVLEIIMAKNENSHA-----FIRKVENIKQTK-----D 1052
Db 1340 ISLKNTALMNSSQOQESHANRSRSIDETLEEVDNESIHEIDQOMKSTQLODKNVANHNSNIC 1399
QY 1053 AOGPDAKMNKEKLYVCVYAMNIIIMSKSTTYSLESKPOVLPARFETOPDKNFSTNKVYL 1112
Db 1400 STKSDEVDYTE-LHESIDIPQSEVNAVOPIEVLTSELKAVTNRSIKTINPDHNVVSONPL 1458
QY 1113 PPEMKSFTTPGKPKT-----TNVLGAVNKPPLSSA----- 1141
Db 1459 KRPSKETPTSEKRSKGHETWVDLVYSEQAVSPSSDYICTNIKSIANESSSLRNSIK 1518
QY 1142 -----GKOSOTKSSRMETVSNASSSSNPSSGRITKGRLDSE----- 1178
Db 1519 VETNCNENSLVTLIDDOQITIKEDKGQVEHVORQEOQESNKKJNSKSTFQDNIAOYKS 1578
QY 1179 ---MDHSEMEDYTMSPLPGKSKDKRDSDLRSE-----LEKPRGKKTPTVEQBE 1227
Db 1579 VAKARPNNEGENNDAICNV-----EQASPVRNEYPGCIGIQIPSTILLNNSKOTE 1628
QY 1228 KLGMDLTK-----LVQEOKPKGORSRKRKGTASEDEQOEMBEKRLKEDILEN--- 1277
Db 1629 KSKVDLNSDEDEHGTVAQEKHQGAINSRNNNDNRMDSTPIQGFEEESREVVMTEEGIN 1688
QY 1278 ---EDQNSPPKKKGRGPRKPLGG-----PKREPTMTKSKGSKSG 1320
Db 1689 VLEDSGTGCELNKNLKG---PLGDKDANINDEVPVEENYRVDGFLKSMHAYSKETG 1744
QY 1321 PPAPEEPEEE-----ROSGNTEOKSKROHVSRRRAQOARSPRESS 1362
Db 1745 ---LEQPEVADISVLPETIRPIPLFNSLKMGSQIKERKLKRLQRELMAPPSPPRM 1799
QY 1363 ALESTQSTPQKGRGRPSKT 1381
Db 1800 T-ENTNINANOGLDTPPKT 1817

RESULT 7
ALM1_SCHPO
ID ALM1_SCHPO STANDARD: PRT: 1727 AA.
AC 09UTK5; 01313; 09UTR8;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Abnormal Long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson K.,
RA Holtroyd S., Hornsby T., Howarth S., Buckle E.J., Hunt S., Jagsels G.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

```

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolckaert G., Aert R., Robben J., Grymoprez B.,
 RA Weltens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
 RA Eber P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R.R., Roeder M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Spakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
 RC STRAIN-972;
 RX MEDLINE-20123449; PubMed-10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The *alm1* gene from Schizosaccharomyces pombe encodes a coiled-coil
 RT protein that associates with the medial region during mitosis.";
 RL Mol. Gen. Genet. 262:921-930(2000).
 RN [3]
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN-968 h90;
 RX MEDLINE-20223868; PubMed-10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
 RL Genes Cells 5:169-190(2000).
 CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: A133357; CAB62414.1; -
 DR EMBL: AF010473; AAB65416.1; ALT_INT.
 DR EMBL: AB028012; BAA87316.1; -
 DR PIR: T50073; T50073.
 DR GenBank: spombe; SPAC1486.04c; -
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 FT SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 3.1%; Score 220.5; DB 1; Length 1727;
 Best Local Similarity 17.8%; Pred. No. 0.011;
 Matches 266; Conservative 251; Mismatches 543; Indels 433; Gaps 61;
 QY 146 ELDSENFITQLYTLESVINNGNOKVHM-----HMVDLMSSTI 185
 DB 130 ELDEQSKIASSEQSLIARDQIEALONENSHGEQVQSAHQALSDIEERKKOHMFSSSSSRV 189
 QY 186 CEGDTVQOE-LLDTVLVNLVPAHKNL-----NKQAVDLAKAL-----LKRTAAQA 228
 DB 190 KEELVQEKALVSDIASLQSDHSGKCEKLEVVSRQVODLEKRLAGLAQONTLEIKIQL 249

QY 229 IEPTITFE-----FNQVLMKQTSISDLSEHV--FDLILEYINIDSHLLSVLPQLEFKL 281
 DB 250 FEQKRSYVSSDGNISKILETDPSTIKLEEEVEQTKRLFLALWESKSELQSEVAALQEKL 309
 QY 282 KS-----NDNEER-----LOYK----- 294
 DB 310 TSQOSLYNNVTEELNNKQOLLSSENSLRELOEKYDSVSESLQVKKENKNTVSAGVGLF 369
 QY 295 --LLAKMEGAKDSLA-----SONKPLMOCVLGRFNIDHPVIRECVKPAFASHCLMNPDL 347
 DB 370 SPLAQKLSAVONPEFSTTKYSDMKLQQ-----KVSSLKIQDLRLTKKFSFC----- 418
 QY 348 AKDILEYIKYASHDPEAIRHDYIVTSVTAKKDILLVNDHLNLFVRETRDKRWYRKE 407
 DB 419 -EYKQKRLIPVKKOORSELIVRNININFLS-----ESLETNNNLTKVQABLLST--KKRQE 471
 QY 408 AMMGLAQIYKKYALQSAAGKDAKQAMIKDLIHI-----YYON-S 448
 DB 472 A-----CYLQLTASRQCSDLSREVICIMAEIDLNETKSRNPATVQVALDEYAQNPS 525
 QY 449 IDDLLEVERIFAQY-----WVPHLETTERKKCLYLATLIDLVANVAKLMKMCQMLR 503
 DB 526 TASETLVNKELANFSIKEAVSKTLEIREKVRAL-----ECQVE-IQKVQVQVQISNAVK 579
 QY 504 HQVADLDLTKOPKPTDASVKAIFSKVAVITRNL-----PDGKAKODPM 547
 DB 580 ENSVTLEEQIKNLESLNKKIKNESILNENLKEMLANSRSSILSHNSAGNIIDMKM 639
 QY 548 KF---TOVLEDEDEKIRKQLEVLPSTQSCQAEQVBE---ITKRLKNPQOPTNPLEMT 601
 DB 640 SIDESTRELEKNYEVYNEMTAIQESLS-KKNQDLSEMAIRKELNSKYQQ----- 691
 QY 602 KELLERLAPVHIDESISALIKQ---VNKSIDGADEDEBQVDPDOAIRGLLELKYLSF 658
 DB 692 QLSDRLTNANNVDEAKKEKELRSINQNIQDISQDO-----RASKAEELIHNSL 746
 QY 659 THPIFSHAETFEESLACLKMDERVAEALQIFKNTGSKI-----EEDFPIARS 708
 DB 747 -----AERLKGELNASKGKD-----LKRQROERLISENDKLAERELMISVS 790
 QY 709 ALPLVLLHKKSKGPPROAKVAIHCHIAIFSKET-----QKQIPEPLAKSL 755
 DB 791 DLQFLWQOOLSDAARKVFE-----SEKESLSLQKLEKSNKMSNDLSLOKSL 842
 QY 756 DPSNLEHLPPLVITGHIALLPDQFAPKMSWATFIVDILNDRLPCKTKTKLWPD 815
 DB 843 EKSGIEY-----SRIRITMLEKOSLSEDNKRL--D 872
 QY 816 EYVSPETMVKIQAIKMVRW-----LLGKNKNHKSQSTSLR-----LTTTLLHSDG 863
 DB 873 NQQWME--IKQELNGVIELEKQRFSTLEAKFTQOKNTVSERALLLESSLSDQSKHTS 930
 QY 864 LTBQGIKSPDMSRLRLAAGSAIYKLAQEPCHIEITILEQYQCALAINDCQVROYFA 923
 DB 931 LESQYNSLRITELQLOAAS-----KLAEEVERVRYKTEYDEVRLOT-----S 971
 QY 924 QKLKSGSLRLPLEVAICALCAKADPVKERRAARQCLVKNIVWREVLKQHAAYSEKL 983
 DB 972 ESLEK-----NHLKITSL-----EQRIVIIDELTASSSLRENNITKD---SETR 1012
 QY 984 LSLPEYVVPYTHLHADPDYKVQDIEQLKDVKECLMFLVLEILMAKNENNS-----HA 1038
 DB 1013 VALLLEENKHLNDELSSH--RNAEKQLEKENDYKQOULLVYTEDLRTREDYKEKELLRHA 1070
 QY 1039 FIRKAVENIKO--TKDAQPDDAKANEKLYTCVDAVANNIIMSKSTTYSLESPKVPVLP-- 1094
 DB 1071 DARSTLOKLRDREYTKALEQVED--LNK-----EIALKAGINSEQPPISEKEDPLQOEY 1122
 QY 1095 -----ARFPTQ---PDKNPSNNTKNYLPPEMKSFPPGPKPTTNVGAANKPLPSAGKQ 1144
 DB 1123 YVLKKQNAMLITQLOSSLNLFNFAETIS--PSPDLSVMKLGSLDLONNHKRISKKEITISQ 1181

QY 1100 QPDKNFSNFKNYLPPEMKSFTEPGKPTTNVIGAVNKPSSAGKOSQTKSSRMETVSNAS 1159
 DB 779 --DKPEEKAKSPKEEVKS---PEKAKS-----PLKADAKAPEKEIPKKEEVK--S 822
 QY 1160 SSNNPSPGRIKRLDSSSEMDHSENDYMTSSPLPKKSKRSDSLVRELKPKGRKK 1219
 DB 823 PVKEEEKPOHVKV---EPPKAEKEKAPATPTEKKSKKE-----EAP--KKE 868
 QY 1220 TPVTEOEKLGMDLTKLVEQPKKGSQSRKRGHTASE---SDEQOMPEEKRLKEDIL 1275
 DB 869 APRPKYEEK-----KEPAVEKPKESKVAKKEEEDKKKVPTEPEKAPAKYEVNEDAK 921
 QY 1276 ENEDQ--NSPPKKGKRGRRPKPLGGTPEKPEPTMTKSKKSGKSGPPAPEEERQ 1333
 DB 922 PKEKTEVAKKEPDDAKAKEPSKP-----AAKKEAPEKKDKTKEEKAK 963
 QY 1334 SGTEBKSQSKQHRVRRNQRAESPSSAIESTOSTPOKGRGRPSKTPSPQPK 1388
 DB 964 KPEKPKTEKAKEDDKTUSKEPSKPAEKAKSSSTDQKSKPPEKATEDKAK 1018

RESULT 9
 USOL YEAST STANDARD; PRT; 1790 AA.

ID USOL YEAST STANDARD; PRT; 1790 AA.
 AC P2536;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 GN USOL OR INT1 OR YD058M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RX MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae.";
 RL J Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YEL047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X54378; CAA38253.1; -
 CC EMBL: L03188; AAB00143.1; -
 CC EMBL: U53668; AAB6659.1; -

DR SGD: S0002216; USOL.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR006955; Usol_p115_C.
 DR InterPro: IPR006953; usol_p115_head.
 DR Pfam: PF04871; Usol_p115_C_1.
 DR Pfam: PF04869; Usol_p115_head_1.
 DR K: P04869; Usol_p115_head_1.
 KW Transport; protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (PORENTAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASF/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G->E (IN REF. 2).
 FT CONFLICT 924 924 E->K (IN REF. 2).
 FT CONFLICT 1253 1253 V->I (IN REF. 2).
 FT CONFLICT 1319 1319 I->V (IN REF. 2).
 FT CONFLICT 1461 1461 N->S (IN REF. 2).
 FT CONFLICT 1581 1581 G->S (IN REF. 2).
 FT CONFLICT 1600 1600 I->V (IN REF. 2).
 FT CONFLICT 1661 1661 R->S (IN REF. 2).
 FT CONFLICT 1772 1772 D->DEEDDEE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 3.08; Score 217; DB 1; Length 1790;
 Best Local Similarity 17.5%; Pred. No. 0.016;
 Matches 256; Conservative 252; Mismatches 520; Indels 433; Gaps 61;

QY 36 KMYVKTMDQDQSEEEKELYINLALIASDFLKHGKDVRLVACGLADIPRIYAPPA 95
 DB 599 KSVIKSLISFSYIODE-----DYIKLVYMLGVAEFSKES 638
 QY 96 PYTSPDKLDMFTFTRQL-----KGLDTPKSPQ- 124
 DB 639 PFPR-----KEYFEFITKTLKDNVYASRIKQFKKDSYFSKYVNMEDSILPDELDETLGPR 694
 QY 125 -ENRYFYLL--ENIAVWKS-----YNCF--ELDSNIFPOLYRTPLFSYINNGH 169
 DB 695 YFSYFPLQLENENIYRRTALSHDPEEPISKISFEVEKLOQCKTGLGEITSL--QTE 752
 QY 170 NOKVHHMVDLMSIICEGPTVSQE-----LDFVLVNLVPAHKLMLNK 213
 DB 753 TESTHETLTKLALNENHEDEKQYILNSSHSLKENSILLETAKNVRSDLDKMT-Q 811
 QY 214 AYDLAKALKRTAQAEPTTFPNQ-----VLMGKTS--ISDLSHEVPD 257
 DB 812 LRVLLETKQENQTALETKSTIHKQEDSITKEGLETLSQKKAEDGINKGRDLFA 871
 QY 258 LIIELVNIDSHLLSVLPQLEFKLSNDNERLOVYKLAKMGARDSELASQNKPL-- 314
 DB 872 LSREMOAVERN-----CKNLQ--REKDSNVNHOKETKSLKEDIAAKITETKALNENLEEM 925
 QY 315 -WQCYLGFENDIHVIRLECFVASHCLMNPDLAKDLTEYLVKRSHPDEE-AIRHDVIV 372
 DB 926 KICG--NNLSKEKEHISKELVEYKSR--FQSHDLVAKLTPEKLSLANNKQMDQAKNESLI 982
 QY 373 SIYTAKKDILLVNDHL-----NEVER--TLDKRWYRKEMAGLAQIYKKA 420
 DB 983 KAVEESKNESIOSLNQKIDMSOEKENFOIERGSIENIKINIBOLKKTISDLQTEKEII 1042
 QY 421 LOSAACKDA-AKQIAMIKKLLHIYVONSIDDLVLRIFAQYVWVHNHETTERMCIVY 479
 DB 1043 SKSDSSKDEYESOISLKEK-----LETA----- 1066
 QY 480 LVYTLIDNAKALNEMWKNLIRHOVKDLDLIDIKPKTDASVAIFSKVMVITRNLDP 539
 DB 1067 --TTADENVNKISELTK-----TRRELALAAVKNLKNLLEK----- 1104
 QY 540 GKADPMKFTQVLEDEKIRKQLEVLSVPTSCQKQEGCVREITKLGPKPOTNPFE 599
 DB 1105 -----LETSEKALKEV-----KENEHLKREKIQLEKEARETRKOQLN 1141
 QY 600 MIFLLERTAPVUID-----TESTSALIKOVYKNSIDSTADDEDEGCVTTDOAIRAGLE 651

```

Db 1142 SLRANLESLEKEHEDLAAQLKKEEQIANKEHQYNEEISQULNDEITSTQOENESJK----- 1197
OY 652 LKVLSTFTHPISTHSAETFEESLACLKMDDEKVAAALQIFKNTSKIEEDPPIHSALL 711
Db 1198 -----KKNDE--LEGEVAKAMST-----SEQSNLKKESEI 1225
OY 712 PVLHHSKKGGPPROAKVAIHCIHAIFSSKETQFAOIFPLIKSLDPSNLEHITPLVITG 771
Db 1226 DALNMQIKE-----LKKNETNEASLESTI-KSVSEIVK----- 1259
OY 772 HIALIAPQFAAPKSWATFIYKDL-MNDRL-----PGKTTKLWVPDEEVSPEYWK 825
Db 1260 -IKELQDE-----CNFEKEVSELEDKASEDKNSKYELQKESKIKRELDA 1308
OY 826 IQATKMWRYMLGKMNHSKSGTSLRLTLTHSDGLTEGKISKDMSRLRLAASA 885
Db 1309 TTEKLITQLEKRTNLSKAKESSELSRLKKT-----SSERRKAEQOLEKLEK----- 1355
OY 886 IVKLAQEPYHEI--ITTEYOYOLCALAINDCYQVQVPAOK-----LHKGLSLRLPLEYMA 941
Db 1356 -----NEIQKNAFEKERKLNEGSSTIQEYSEKINTLEDELIRLQNE-- 1402
OY 942 ICALCANDPKVERPAHARQCLVKNNVREYLKQAAVSEKLLSLPEYVVPYTHLLAH 1001
Db 1403 ---LKAKE-IDNTRSELEKVSLSNDELLEE--KQNTIKS-----LODEILSYKDKITRN 1450
OY 1002 DPDVVKVQ-----DIEOLKDYKECLMFEYLEILMAKENNSNAIFIKKVENIKQKDAQCP 1056
Db 1451 DEKLISTERDNKRDLSELK-----EQLRRAQESKA-----KVEEGKLKLEESSEK 1495
OY 1057 DDAKANEKLYTVCDVAMNIIKSKSTTYLSESPKDPVLPAFPPTQDKNFSMTKMLPPEM 1116
Db 1496 EKAL-EKSKEMMKKLESTISNET--ELKSSMETI-----RKSDLEBOSKSAEDTI 1546
OY 1117 KSFTPGKPTTNVLGAVN-----KPISSAGKOSQTSKSMETVSNASSS-----NP 1164
Db 1547 KNL-----QHEKSDISLRINESEKDIIEELKSKLRIFAKSGSELETYKQELNNAQEKIRINA 1602
OY 1165 SSPRIGRGLSSSEMDHSENDYTMSSP-----LPGRKSDKRDSDLVRLSELEKRGKK 1219
Db 1603 EENTVLKSKLDIERELKDKQAEIKSNOEEKLLTSRLKELEQELDSTQQAOKKSEEBR 1662
OY 1220 TPYTE-OEEKIGMDLTKLVQ-----DQPKGSGRSRRKRGHTASEDEQOPEEKR 1269
Db 1663 AEVARKFYEKSQDDEKAMLETKYNDLVNKRQAMKRDDYAK-----TTTSQKQELTKL 1717
OY 1270 LKEDILENEQNSPPKKGRGPRPKPLGGTPEEPTM-----KTSKKSGSKK-- 1318
Db 1718 AKE--LDNLKAENSKLKEANEDR-----SEIDMLMLVTDLDEKNAKYRSKLDLG 1766
OY 1319 -----SGPPAPEEEEEEEROSG 1335
Db 1767 VEISSDEDEDEDEDEEERG 1787

```

```

RESULT 10
ANCL_CAEEL
ID ANCL_CAEEL STANDARD: PRT: 8545 AA.
AC Q94M4; O61841; O61842;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nuclear anchorage protein 1 (Anchorage 1 protein) (Nesprin homolog).
CN ANC-1 OR ZK973.6 OR T03A1.3/T03A1.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN SEQUENCE OF 1-1751 FROM N.A., FUNCTION, CHARACTERIZATION, AND
RP INTERACTION WITH F-ACTIN AND UNC-84.
RC STRAIN-Bristol N2;
RX MEDLINE=22264052; PubMed=12169658;

```

```

RA Starr D.A., Han M.;
RT "Role of ANC-1 in tethering nuclei to the actin cytoskeleton.";
RL Science 298:406-409(2002).
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Gattung S., Goela D., Broly M.;
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (Nov-2002) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 689-1243 FROM N.A.
RX MEDLINE=22296983; PubMed=12408964;
RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT "The nesprins are giant actin-binding proteins, orthologous to
RT Drosophila melanogaster muscle protein MSP-300.";
RL Genomics 80:473-481(2002).
CC -1- FUNCTION: Plays a central role in nuclear and mitochondrial
CC anchoring. Probably connects nuclei to the cytoskeleton by
CC interacting with unc-84 at the nuclear envelope and with F-actin
CC in the cytoplasm, creating a bridge across the nuclear envelope
CC between the cytoskeleton and the nucleus.
CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
CC Interacts indirectly with unc-84.
CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (potential). The
CC largest part of the protein is cytoplasmic, while its C-terminal
CC part is associated either with the nuclear envelope, most probably
CC the outer nuclear membrane, or with mitochondrial membrane.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in all postembryonic
CC cells.
CC -1- DEVELOPMENTAL STAGE: First expressed in L1 larvae, and thereafter.
CC -1- DOMAIN: The large coiled coil domains are composed of 6 nearly
CC exact repeats of 903 residues. The last repeat is partial. These
CC repeats are conserved in Hawai (CB4856), Australia (AB4) and
CC Germany (RC301) strains. The length of the repeat may be
CC maintained because of a selective advantage to keep the protein
CC large and allow a single molecule to extent more than 0.5
CC micrometers.
CC -1- DOMAIN: The Klargicht domain, which contains a potential
CC transmembrane domain, is essential for the nuclear envelope
CC targeting.
CC -1- SIMILARITY: Belongs to the Nesprin family.
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 1 Klargicht domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL, AY157938; AAN35200.1; -
CC EMBL, AC006834; AAF40010.3; -
CC EMBL, AY126454; AAM95163.1; -
CC HSSP, P46939; 10AG.
CC WormPep: ZK973.6; CE33588.
CC InterPro: IPR001589; Actbind_actln.
CC InterPro: IPR001715; Calponin-like.
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00307; CH; 2.
CC SMART: SM00264; BAG; 6.
CC SMART: SM00033; CH; 2.
CC PROSITE: PS00019; ACTININ_1; 1.
CC PROSITE: PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE: PS50021; CH; 2.
CC Structural protein; Cytoskeleton; Actin-binding; Transmembrane;
CC Coiled coil; Repeat.
CC DOMAIN 1 8494 CYTOPLASMIC (POTENTIAL).
FT

```

FT TRANSMEM 8495 8513 ANCHOR FOR TYPE IV MEMBRANE PROTEIN
 FT DOMAIN 8514 8545 (POTENTIAL).
 FT DOMAIN 1 325 PERINUCLEAR SPACE (POTENTIAL).
 FT DOMAIN 23 130 ACTIN-BINDING.
 FT DOMAIN 222 325 CH 1.
 FT REPEAT 3241 8199 CH 2.
 FT REPEAT 4144 5097 6 X TANDEM REPEAT.
 FT REPEAT 5098 6000 1.
 FT REPEAT 6001 6903 2.
 FT REPEAT 6904 7806 3.
 FT REPEAT 7807 8199 4.
 FT DOMAIN 754 774 5.
 FT DOMAIN 1072 1101 6.
 FT DOMAIN 1215 1236 COILED COIL (POTENTIAL).
 FT DOMAIN 1324 1384 COILED COIL (POTENTIAL).
 FT DOMAIN 1574 1629 COILED COIL (POTENTIAL).
 FT DOMAIN 1725 1754 COILED COIL (POTENTIAL).
 FT DOMAIN 1950 1981 COILED COIL (POTENTIAL).
 FT DOMAIN 2103 2580 COILED COIL (POTENTIAL).
 FT DOMAIN 2682 2712 COILED COIL (POTENTIAL).
 FT DOMAIN 2852 2949 COILED COIL (POTENTIAL).
 FT DOMAIN 3002 3119 COILED COIL (POTENTIAL).
 FT DOMAIN 3178 3295 COILED COIL (POTENTIAL).
 FT DOMAIN 3346 3417 COILED COIL (POTENTIAL).
 FT DOMAIN 3482 3552 COILED COIL (POTENTIAL).
 FT DOMAIN 3587 3703 COILED COIL (POTENTIAL).
 FT DOMAIN 3781 3839 COILED COIL (POTENTIAL).
 FT DOMAIN 3902 4022 COILED COIL (POTENTIAL).
 FT DOMAIN 4114 4198 COILED COIL (POTENTIAL).
 FT DOMAIN 4249 4320 COILED COIL (POTENTIAL).
 FT DOMAIN 4436 4506 COILED COIL (POTENTIAL).
 FT DOMAIN 4541 4657 COILED COIL (POTENTIAL).
 FT DOMAIN 4735 4793 COILED COIL (POTENTIAL).
 FT DOMAIN 4856 4976 COILED COIL (POTENTIAL).
 FT DOMAIN 5035 5152 COILED COIL (POTENTIAL).
 FT DOMAIN 5203 5274 COILED COIL (POTENTIAL).
 FT DOMAIN 5339 5409 COILED COIL (POTENTIAL).
 FT DOMAIN 5444 5560 COILED COIL (POTENTIAL).
 FT DOMAIN 5638 5696 COILED COIL (POTENTIAL).
 FT DOMAIN 5759 5879 COILED COIL (POTENTIAL).
 FT DOMAIN 5938 6055 COILED COIL (POTENTIAL).
 FT DOMAIN 6106 6177 COILED COIL (POTENTIAL).
 FT DOMAIN 6242 6312 COILED COIL (POTENTIAL).
 FT DOMAIN 6347 6463 COILED COIL (POTENTIAL).
 FT DOMAIN 6541 6599 COILED COIL (POTENTIAL).
 FT DOMAIN 6662 6782 COILED COIL (POTENTIAL).
 FT DOMAIN 6841 6958 COILED COIL (POTENTIAL).
 FT DOMAIN 7009 7080 COILED COIL (POTENTIAL).
 FT DOMAIN 7145 7215 COILED COIL (POTENTIAL).
 FT DOMAIN 7250 7366 COILED COIL (POTENTIAL).
 FT DOMAIN 7444 7502 COILED COIL (POTENTIAL).
 FT DOMAIN 7565 7685 COILED COIL (POTENTIAL).
 FT DOMAIN 7744 7861 COILED COIL (POTENTIAL).
 FT DOMAIN 7912 7983 COILED COIL (POTENTIAL).
 FT DOMAIN 8048 8118 COILED COIL (POTENTIAL).
 FT DOMAIN 8153 8204 COILED COIL (POTENTIAL).
 FT DOMAIN 8273 8329 COILED COIL (POTENTIAL).
 FT DOMAIN 8370 8390 COILED COIL (POTENTIAL).
 FT DOMAIN 8486 8545 KLARSICHT.
 FT DOMAIN 1655 1658 POLY-ARG.
 FT DOMAIN 3021 3029 POLY-LYS.
 FT DOMAIN 3924 3932 POLY-LYS.
 FT DOMAIN 4878 4886 POLY-LYS.
 FT DOMAIN 5781 5789 POLY-LYS.
 FT DOMAIN 6684 6692 POLY-LYS.
 FT DOMAIN 7587 7595 POLY-LYS.
 SQ SEQUENCE 8545 AA; 956470 MM; 02A94D94BEE19E3 CRC64;

QY 19 VKEISDKIS-KEEMVRLKAVYKTFMD---MDQSEEEKELYLNALHL-----ASDF 67
 DB 5221 VNEIARKLQVSOQLSVQEVPEASLDEQKQLEEDVENQKNHNLNLANLPENDPTADEL 5280
 QY 68 FLKHGPG---KDVRLVACCLADIRIYAPAPYTSPOKLDIFMTROKLGEDTS 122
 DB 5281 ROKSQWDLRSRLKRLQQLQSAVGD--KLALAAFNARRNABDALDITRE--DGGDNKS 5337
 QY 123 P-----QFNRYVLEENIATWKSYNICELEDS-----NEIFPOLYRLFYSINNG 168
 DB 5338 PDELIDDLAKKETTYAKLIDTVYSGVP---DELDEKERAEYNDLARL-ATRADVLKN- 5391
 QY 169 HNOQVHMVMDLSSIIICGDTVS--OELDTVLVNLVPAKMLNKQAYDLAKAL----- 221
 DB 5392 -----KRAELQAVAKADEKSLHSDVRIYSRLVPLVREDELRLH--AEAVPQYAV 5443
 QY 222 ----LKRTQAIEPYITTF-----ENQVIMLKTSISDSSEHVFDLILEYN---- 264
 DB 5444 KAEELKKEVAAKAVIANAPSSDAHVOLEQAVATAETLIPLEERA-----RLMNEFLA 5498
 QY 265 ----IDSHLLSVLPQLEFKLS-----NDNEERLOVVKIL---AKMEGAKDSEIASON 311
 DB 5499 ARNDIDP-----ALIBOLOQPLDAVLAPKRSABEAAQDVENLRNNSQOQLSDLDKNTANQ 5553
 QY 312 K-----PLWQCYLG--RF-----NDIHVPIRECV--KFASHCLMNHDP 346
 DB 5554 RISELDDPLESAVADVFEDVDAQTRHQYDVLNDVNAELDEDETLKQSAQVANEIDD 5613
 QY 347 LAMDLELYKVRHDEEAI RHVIVYSIYTAAKKDLLVNDHLNFRER---TLCKRR 403
 DB 5614 ISK-----MIDSTDERSTLDITRIKSDIPALQAQINRIKRIVNDASRKHTVTPKTA 5667
 QY 404 VRKEAMGIAQIYKKKALOSAGKDAKQIAWKIDLHLYYONSI-----DBRLAYE 456
 DB 5668 EDLDNKLAKQTLDAIKITSDEHDKQQLLSLK--LINSQEQIPLDQKSDDLKTPAE 5724
 QY 457 RIFAQYVPHNLEETTERMKCLYYLVATFLDINAVKALNEMKKCONLLRHQVKDIL----- 510
 DB 5725 KEITNSLKPEAEPL-----LAKIOELREAKRVDEARSAAHQDIVALERE 5770
 QY 511 -----DLIQPK---IDASKALFESKVM-----ITRLPRLPGKAOPE 5825
 DB 5771 AEDVTAKESAKKKKKKKKSPQEMIDELSAKVEAKALIPKIEBAKNEPLP-----AD 5825
 QY 546 MKRFTQVLEDEDEKIRKQLEVLVSPSCSKQAE-----GCVEITKRGKGNPKQ-----PT 594
 DB 5826 KPKAEQVLSNLEAFVAVDETQVSE-----KODELDKLNANDAIKRLGDALDADEKTVVES 5881
 QY 595 N-PLEMIKIFLERIAPVHIDT-----BSISALIKQVKSIDGTA 633
 DB 5882 SVPALESEFK---DRIAP--HIAITLVEAVNDVPASVEPSAVALRBRARAFVSDLEKKNIOKGTG 5937
 QY 634 DDE-----DEGVTTQDA-----IRAGLELAKYLSTHET 662
 DB 5938 DDEKRADELKNDGNAVKNVEDVSKYQNDQPLDVAKADANKLKTATVQDLTKLA----- 5992
 QY 663 SFPSAEFTFESILACLKMDDEKVAEALQ-----IFKNGSKIEEDFPHRSALP 712
 DB 5993 --ESSQKIDPQVAKDLKDKSTAKKELLQALEKAIPOBDALRRQADALNRLNLEKELTK 6050
 QY 713 VLHHKSKKGPPOAKYAIHCIAHIFSSKETQFAQIEPRLKSLDPSNLEHLITPLVYIGH 772
 DB 6051 VDEFKRED-----ALPIVDQLANTMTL-----KRTDSSNNEKAVAPSSLSH 6093
 QY 773 IALLADQFAAPKSNVAFIYVD-----LLMDRLPGRKTTTLAMPDEE---VSPR 821
 DB 6094 DDLIVV---GLPEKVPOLQHAIDDKQALNKAANAVEIAP---KLQVSOQLQSVQPE 6144
 QY 822 TWVKIQAIKMVMKWLGMKNHNSKSGSTLRLTLTILHSGDLEEGKISKSPMSRLR-- 879
 DB 6145 VPASLDEQKQLEDEVENQKNHLE-----NLLANLPENDPTADELRKQSQWDLRSRLKDL 6197

Query Match 3.0%; Score 212.5; DB 1; Length 8545;
 Best Local Similarity 18.4%; Pred. No. 0.2; 642; Indels 495; Gaps 79;
 Matches 317; Conservative 268; Mismatches 642; Indels 495; Gaps 79;

```

OY 880 -----LAAGSAIVKLAQEPCCY-----HEITLEQYCALAI 911
DB 6198 LKQGSAGVDKIALAALAAFNAAKNMEDALDITREDGGDDKSPBELI----- 6245
OY 912 NDECYQVROVPAQKLIH--KGLSRLRP-----LEYMAICA--LCARDPVKERRAHARQCLV 963
DB 6246 -DLAKKEETVAKLIDTJVSQVPELDLDDKERAEYNDLLARLATADVLKKNRAELEQA-- 6302
OY 964 KNINVRREYIKQHAANVSEKLSLIP-----EYV-----VPYTIH 997
DB 6303 --VAKADEKSLHDSVDRIVSRLVPSREDELRHNAEVPQYAPKAELKEVEAKA 6360
OY 998 LLAHDPDY-VKVDIOEIO-----LKDYKECLMFWLETLAKKNENNSHAFIRKKNVENIX 1048
DB 6361 VIANPSSDAHVQOLEQAVATAETILIPDLERARIMNEFLARND-----IDALIEQIQ 6414
OY 1049 QTKDA-----QGPDDAKMNEKLYTVCVAMNIMSKSTYSLESPKDPVLPARF 1097
DB 6415 QPLDAVLAQPKRSAAEAQDVENLNNSQQLSDLNKIANLQRISELDPLESAVADRF 6474
OY 1098 F-----TOPDKNFSNTKNTLPPEM--KSEFFTECKPPTTVNLGAVNKLSSAGKQOTKSSR 1151
DB 6475 FDVAEQTROHQYDVLNDVNAELEDEETLKOSASQVANEIDISKMIDS---TDPERSI 6530
OY 1152 MEVSNASSSSNPSSPGRIKGR-----DSEMDHSENEEDYTMSSPLPG--KK 1197
DB 6531 LDTIAKSDIPALKQAINIKRIVADASRKHVTTDPKIAEDLDKLAQLQELDLDAIKT 6590
OY 1198 SDRKDDSDLVSELEKPRGRKTPVTEQ---EKLGMDDT---TKLVQEKPGSQ-- 1247
DB 6591 SDEHDKEOILIS-----LKLNISQFEQIPLDQKDDLTKAEKITNSLKPEAEPL 6642
OY 1248 -----RSRRK-GHTASESDQEQWPEEKRLKEDILLENEDQNSPPKGRPRPKPLG 1299
DB 6643 LAKIOELREAKRVGDEARSAHDDIVALEKEAEDVTAKESAKK--KKDKKSPQEMIDE 6700
OY 1300 GTPK--EEPTMKTSKSGSKSGPPAPEEEEEEQSGNTEQSKSKOHVSRRAQOQRAE 1357
DB 6701 LSAKVVEAKKALIPKEEAKNENLPAADKPKAEQLVS-NLEAFVADVEQVSEKQDELDK 6759
OY 1358 SPSS-----AIESQSTPPQKGRPKTPSPSQPKKNV 1391
DB 6760 LNNANDAIKRLGDLDAEKTIV-----PSSVPALSEKDR 6796

```

```

RESULT 11
NFH_RAT STANDARD: PRT: 831 AA.
ID NFH_RAT
AC P16884; Q63368;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
GN NFH OR NFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143606;
RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites."
RL FEBS Lett. 241:213-218(1988).
RN [2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and

```

```

RT in situ detection."
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN [3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderson B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H)."
RL FEBS Lett. 209:203-205(1986).
RN [4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Bretner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M37227; AAA41693.1; ALT_FRAME.
DR EMBL: X13804; CA932038.1; ALT_FRAME.
DR EMBL: M21964; AAA41695.1; -.
DR EMBL: J04517; AAA41692.1; -.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; Filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neutrone; phosphorylation; Repeat.
FT NON_TER 1
FT DOMAIN 1
FT 276 641 51 x 3 AA TANDEN REPEATS OF K-S-P.
FT 164 164 L->I (IN REF. 2).
FT 185 185 I->S (IN REF. 2).
FT 193 193 L->T (IN REF. 2).
FT 199 199 M->T (IN REF. 2).
FT 346 346 K->N (IN REF. 1).
FT 373 373 A->V (IN REF. 1 AND 4).
FT 482 482 G->E (IN REF. 2 AND 4).
FT 485 485 P->S (IN REF. 2).
FT 570 571 RK->KE (IN REF. 2 AND 4).
FT 591 591 P->T (IN REF. 2 AND 4).
FT 727 727 A->V (IN REF. 4).
FT 757 759 AAP->GST (IN REF. 4).
FT 769 769 T->L (IN REF. 2).
FT 775 775 R->P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973Cf313EE768 CRC64;

```

```

Query Match 2.9%; Score 212; DB 1; Length 831;
Best Local Similarity 22.9%; Pred. No. 0.01;
Matches 85; Conservative 56; Mismatches 150; Indels 80; Gaps 13;

```



```

OY 445 -YONSI-----DDRLVERIFAQVWVPHNLETTERMKCLYIATL-----DIANVK 490
DB 4163 DIOSALSVSTJOEADSRLLTE-----ISSWDGQOKFYRSIARHARLMAAL 4210
OY 491 A-----LNMKKCONLLRHQVLDLILKOPKTDASKAFSKVMVTRNL-----536
DB 4211 ATRPAKMGKGNVERGRFSAHLMK---MLVRQRBSITTL-----SEOWILIRNLSCVOET 4263
OY 537 -----PDGKAODEMKKFTVOLEDEKIRKQLELVVS-----PTSCQKQAE 577
DB 4264 HSRIMGPOAYVPAFPPDOGVQOWTERL---QHLAMQOQLLBQSLMLCCSVGAPRGH 4320
OY 578 GCVRITTKKLGKPKOP-----TNPLEMIKFLLENIAPV-----611
DB 4321 GNV-----GVLGQPPGCLLEGPELSKQQLCGVLDLPSNLVSPSPGSQLPSGCMRKO 4376
OY 612 -HIDTESIALIKQVNSIDGTADEDEGVPTDOAIRAGLELLKVLSPFHSPHSARET 670
DB 4377 DHLMOGSTTRL-----TEMLKTIKIVKADVD---KIROQSCETLPHSMKPE 4419
OY 671 F---SLACL-----KMODEKVAEALQIFKWTGSKIEEDPH 705
DB 4420 EVCSSALSQSVNHLQSLFTLPGMEVEQOROSOMALVSELEVREGEISKAMADFTT 4479
OY 706 IRSALL-----PYLHHSKKGPPROAKVAIHICHAIFSSKEQFOAIFEPKHSIDP 757
DB 4480 KTHLLTSDSGONGMDEGEVEDSEQMEIARILCAIQULEKKNKKALENTQASP 4539
OY 758 ---SNLEHLITPLVTIGHIALLAPDOFAAPKSVAFIVDLMNDRLPGKTKTKLV 813
DB 4540 QEDYAGFERLQS-----GHILTKLEDDF-----WA 4564
OY 814 PDEEVSPTMWKIQAIKMMVRWLLGMKNHKSQSTYLRL-----LTTILHSDG 862
DB 4565 ---DVS---TLHVQITISAISELLERLKSQYGEDGTAKLFPQSOSCLVRLVLPYLSYS 4618
OY 863 DLTEQGISKPDMSRLAAGSAIYKLAQEPCHYHITLEVOQLCALAINDCYQVROFE 922
DB 4619 DLV-----LEFLTMSLATHRSTAKL-----LSVLA-----QVF 4646
OY 923 AOKLHKGSLRLPLEYMAICALKADPKERRAHAROLVKNINVRREYLKOHAAVSEK 982
DB 4647 TELAQKGRFC---LPREFM-----EDSAGEGATE-----4671
OY 983 LLSLPEYVPTTILHLADHPYVYVQDIEOLKDYKECLMFWLEILMAKNENNSHAFTRK 1042
DB 4672 -----FHDEGGGIGEGEGMKDVSD-----QIGNEQOVEDTQK 4705
OY 1043 MYENIKOTKDA---QGPDDA-KMNE---KLITYCDVAMNTIMSKSTY-----S 1084
DB 4706 GOEKKEPDPSKSDIKGEDNALIKEMSEDDGKMHODELEBOEEDDEKSDSEGDLKHMGD 4765
OY 1085 LESPRDVLPAFRTQDPKNSNTKNYLPRPKKSFTT-PGKRTYNVLGAVKPLPSAGK 1143
DB 4766 LNGEADKLDERLMDDEDEDEDEE---DNKTETGQMGDEEDSELVAKDNLDSDS-CN 4820
OY 1144 OSQOTS---SRMETVSNASSSSNPSPGRIKGLDSEMDHSHENEDYTMS---PLPGKKS 1198
DB 4821 SNKDSQODKKEKEEAEADOGGOGEDKINEDIDRDYDENENVDRHGOEKVPRP-EAL 4879
OY 1199 DKRDSDLVSELEKPRGRKRTPYVEOEKLCMDLTLKYOOROKSGRSRKSHT---1255
DB 4880 DLPRDLNL---DSEKNGEOTDNEGEENPLE-----IKKPEAGHAEKEKETELD 4931
OY 1256 ASESDQOWPEERKLKEDILENEDQON---SPPKKGRGRPRPKPLGGTGRPEPTTKTSK 1313
DB 4932 QNESQSPQEPGESEDDKAKGEEMGTGADQDQDAQNHPE---HSPEQOQSVEEKDK 4988
OY 1314 GSKKSGPPAP-----EEEEEROSGNTBOK---SKSKHRSRRRAQRAESPSS-A 1363
DB 4989 EADEGGENGADGQFOPEEEREDSDTEEOVPALEERKEH---ASCQGTGVENMQNTQA 5046

```

```

OY 1364 IESTOSTPOKGRGR 1377
DB 5047 MELAGAPKEQOK 5060

RESULT 13
BPAL_MOUSE
ID BPAL_MOUSE STANDARD: PRT: 7389 AA.
AC Q91206; Q91207;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1, isoforms 1/2/3/4 (BPA) (Hemidesmosomal
DE plaque protein) (Dystonia musculorum protein) (Dystonin).
GN BPA1 OR DST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RX STRAIN-BALB/c; TISSUE=Muscle, and Neuron.
RC MEDLINE=21405767; PubMed=11514586;
RA Leung C.L., Zheng M., Prater S.M., Liem R.K.H.;
RT "The BPA1 locus: alternative splicing produces multiple isoforms with
RT distinct cytoskeletal linker domains, including predominant isoforms
RT in neurons and muscles."
RL J. Cell Biol. 154:691-697(2001).
RN [2]
RP SEQUENCE OF 6693-7389 FROM N.A. (ISOFORMS 3 AND 4).
RX STRAIN-C57BL/6J; TISSUE=Fetal skin, and Fetal spinal cord;
RC MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nigaki I., Osato N., Saito R., Suzuki H., Yamataka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schiml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Gilmour S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kaniel A., Kawai H., Kawasawa Y., Kedzielski R.M., King B.L.,
RA Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numa K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verrado R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilting L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carlini P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [3]
RP FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
RP intermediate filaments to the inner plaque of hemidesmosomes. The
RP proteins may self-aggregate to form filaments or a two-dimensional
RP mesh (By similarity).
CC -!- SUBUNIT: Homodimer. Interacts with the neuronal intermediate
CC filament protein, Prph (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=7;
CC Name=2; Synonyms=b;
CC IsoId=Q91206-1; Sequence=Displayed;
CC Name=1; Synonyms=a;

```

CC IsoId=Q91ZU6-2; Sequence=VSP_050483;
 CC Name=3;
 CC IsoId=Q91ZU6-3; Sequence=VSP_050484, VSP_050485, VSP_050486;
 CC Note=No experimental confirmation available;
 CC Name=4;
 CC IsoId=Q91ZU6-4; Sequence=VSP_050485, VSP_050486;
 CC Note=No experimental confirmation available;
 CC Name=5; Synonyms=;
 CC IsoId=Q91ZU8-1; Sequence=External;
 CC Name=6; Synonyms=nl;
 CC IsoId=Q60824-1; Sequence=External;
 CC Name=7; Synonyms=nl;
 CC IsoId=Q60824-2; Sequence=External;
 CC -1- TISSUE SPECIFICITY: Expressed at high levels in the heart and skeletal muscle and at low levels in the skin in the adult.
 CC Expressed in the myocardium, skeletal muscle masses, vertebrae cartilage, and epithelia of the tongue of 14.5 day embryos.
 CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -1- SIMILARITY: Contains 1 actin-binding domain.
 CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 9 plectrin repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF396879; AAK83384.1; -;
 DR EMBL: AF396878; AAK83383.1; -;
 DR EMBL: AK051626; BAC34695.1; -;
 DR EMBL: AK037206; BAC29753.1; -;
 DR MGD: MGI:104627; Dst.
 DR GO: GO:0005737; C:cytoplasm; ISS.
 DR GO: GO:0030056; C:hemidesmosome; IDA.
 DR GO: GO:0005200; F:structural constituent of cytoskeleton; ISS.
 DR GO: GO:0045104; P:intermediate filament cytoskeleton organiza. . .; ISS.
 DR InterPro: IPR001589; Actbind_actinin.
 DR InterPro: IPR001715; Calponin-like.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003108; GAS2.
 DR InterPro: IPR001101; Plectrin_repeat.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00036; ehand; 2.
 DR Pfam: PF02187; GAS2; 1.
 DR Pfam: PF00681; Plectrin; 5.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00435; spectrin; 26.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00053; CH; 2.
 DR SMART: SM00054; Efn; 2.
 DR SMART: SM00243; GAS2; 1.
 DR SMART: SM00250; PLEC; 9.
 DR SMART: SM00150; SPC; 34.
 DR PROSITE: PS00019; ACTININ_1; 1.
 DR PROSITE: PS50021; CH; 2.
 DR PROSITE: PS00018; EF_HAND; 2.
 KM Actin-binding; Coiled coil; Repeat: SH3 domain; Structural protein;
 KM Cytoskeleton; Cell adhesion; Calcium; Calcium-binding;
 KM Alternative splicing.
 FT DOMAIN 31 255 ACTIN-BINDING.
 FT DOMAIN 35 138 CH 1.
 FT DOMAIN 151 252 CH 2.
 FT REPEAT 590 667 SPECTRIN 1.
 FT REPEAT 675 770 SPECTRIN 2.
 FT DOMAIN 889 941 SH3.

FT REPEAT 1260 1356 SPECTRIN 3.
 FT REPEAT 1537 1581 PLECTIN 1.
 FT REPEAT 1619 1657 PLECTIN 2.
 FT REPEAT 1657 1694 PLECTIN 3.
 FT REPEAT 1695 1732 PLECTIN 4.
 FT REPEAT 1735 1770 PLECTIN 5.
 FT REPEAT 1771 1808 PLECTIN 6.
 FT REPEAT 1811 1846 PLECTIN 7.
 FT REPEAT 1847 1884 PLECTIN 8.
 FT REPEAT 1886 1922 PLECTIN 9.
 FT REPEAT 3814 3914 SPECTRIN 4.
 FT REPEAT 4053 4152 SPECTRIN 5.
 FT REPEAT 4270 4346 SPECTRIN 6.
 FT REPEAT 4409 4517 SPECTRIN 7.
 FT REPEAT 4522 4620 SPECTRIN 8.
 FT REPEAT 4623 4729 SPECTRIN 9.
 FT REPEAT 4742 4840 SPECTRIN 10.
 FT REPEAT 4851 4949 SPECTRIN 11.
 FT REPEAT 5177 5278 SPECTRIN 12.
 FT REPEAT 5288 5385 SPECTRIN 13.
 FT REPEAT 5397 5497 SPECTRIN 14.
 FT REPEAT 5506 5605 SPECTRIN 15.
 FT REPEAT 5646 5714 SPECTRIN 16.
 FT REPEAT 5725 5824 SPECTRIN 17.
 FT REPEAT 5946 6046 SPECTRIN 18.
 FT REPEAT 6055 6155 SPECTRIN 19.
 FT REPEAT 6165 6265 SPECTRIN 20.
 FT REPEAT 6274 6372 SPECTRIN 21.
 FT REPEAT 6383 6480 SPECTRIN 22.
 FT REPEAT 6492 6592 SPECTRIN 23.
 FT REPEAT 6602 6701 SPECTRIN 24.
 FT REPEAT 6710 6808 SPECTRIN 25.
 FT REPEAT 6826 6914 SPECTRIN 26.
 FT REPEAT 6962 7020 SPECTRIN 27.
 FT REPEAT 7028 7040 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 7064 7076 EF-HAND 2 (POTENTIAL).
 FT CA_BIND 7064 7076 COILED COIL (POTENTIAL).
 FT DOMAIN 452 486 COILED COIL (POTENTIAL).
 FT DOMAIN 638 703 COILED COIL (POTENTIAL).
 FT DOMAIN 737 773 COILED COIL (POTENTIAL).
 FT DOMAIN 1003 1138 COILED COIL (POTENTIAL).
 FT DOMAIN 1195 1247 COILED COIL (POTENTIAL).
 FT DOMAIN 1413 1455 COILED COIL (POTENTIAL).
 FT DOMAIN 1504 1527 COILED COIL (POTENTIAL).
 FT DOMAIN 3336 3359 COILED COIL (POTENTIAL).
 FT DOMAIN 3539 3715 COILED COIL (POTENTIAL).
 FT DOMAIN 3809 3893 COILED COIL (POTENTIAL).
 FT DOMAIN 3957 3978 COILED COIL (POTENTIAL).
 FT DOMAIN 4006 4043 COILED COIL (POTENTIAL).
 FT DOMAIN 4159 4209 COILED COIL (POTENTIAL).
 FT DOMAIN 4270 4311 COILED COIL (POTENTIAL).
 FT DOMAIN 4426 4563 COILED COIL (POTENTIAL).
 FT DOMAIN 4696 4735 COILED COIL (POTENTIAL).
 FT DOMAIN 4847 5097 COILED COIL (POTENTIAL).
 FT DOMAIN 5173 5233 COILED COIL (POTENTIAL).
 FT DOMAIN 5570 5603 COILED COIL (POTENTIAL).
 FT DOMAIN 5717 5739 COILED COIL (POTENTIAL).
 FT DOMAIN 5787 5809 COILED COIL (POTENTIAL).
 FT DOMAIN 6010 6089 COILED COIL (POTENTIAL).
 FT DOMAIN 6116 6164 COILED COIL (POTENTIAL).
 FT DOMAIN 6277 6318 COILED COIL (POTENTIAL).
 FT DOMAIN 6381 6417 COILED COIL (POTENTIAL).

Query Match 2.9%; Score 209; DB 1; Length 7389;
 Best Local Similarity 18.2%; Pred. No. 0.24;
 Matches 320; Conservative 247; Mismatches 591; Indels 600; Gaps 77;
 OY 17 PVKE-----ISDKISKEWYRLKMYVKTFMDQOSEEKELYLNLALHLAS 65
 Db 4717 PKIKAQEVCEEDLSALVKEEYLRKAELSRQLEGILSKRDIKTE-----NHVCHLOS 4769
 OY 66 DFPLKHGKXVRLVNAACLDIRIVAPAPYSPDKLKDIFMFIHQJLGLDFTKSPDF 125
 Db 4770 -----ACASSHQFO-----QMSKDFQAWL--DAKKEQRDSPP 4801


```

QY 867 --OGKISKPD-----MSRLRLAAGSA-----IVKLAOEPCHYEITILEOY---- 904
DB 363 LLYGWSSKKNLMEIYKLTMTHTVDKAGETTYRDELTLKIIDIOSQSYQITNFWETISL 422
QY 905 -----QCALAINDECYQROY--FAQKIHKLSRLPLEYMAICALCAKDP 950
DB 423 VELTRLEGRHGLLAOMLDVAIRKAIKRFKFA-----VSOMASLIDSAHLA----- 470
QY 951 VKERRAHAROCIVKNINVRREYLKQAAVSEKLSILPEYVYPTIHLIANDPDVYKOD 1010
DB 471 SSTORNGICEVLYAAAMIGEFSEHLOEPHTLEALRPVTLTLEGHIDA-----VYVON 525
QY 1011 ISOLADVKECLWFEVLEILMAKNE-----NNSHAFIRKMNENIQTKDAQCPDDAKMEKLY 1066
DB 526 VVKL-----YASIIQQKQAGAEAGAAVTOIWMVRLQFQVOSA---DLEVQERAS 573
QY 1067 TVCDVAMNIIIMSKSTTYSLESPKRDYLPARFPTOPDKNFSNTRKNTLPPEMKSEFTPGKPK 1126
DB 574 CILQIVKHI-----QKIQAKDVPA-----EEVSALFAGELNVPAPKAKR--KVYPVE 619
QY 1127 TTNVLGAVNKPPLSSAGKOSQTSRMEVYSNASSSSNP-----SSPGRIKGRIDSSEND 1180
DB 620 GUDLDAMINEPLS-----DSESEDERPRAVFEHEEORRPRKHPSEADEELARRRERKOE 675
QY 1181 HSENDYTWSPPLPKGKSDKR---DDSDLYRSELEKPRGRKKTPTYEQ-----EKLGM 1232
DB 676 QANNFYIKSSPSPOKRRQDTPGVHEIPIVQIDLSVPLKVPGLPDSODVYKLEERRNHQ 735
QY 1233 DITKIVQOKPKGSPRSKRKHTA---SESDEQOMPEEK--RIKEDILEN---EDQNSPP 1285
DB 736 KLEKRRRRKREKREKRRHSLSPTSEDEDIAAQOVDIYTEMPEALPSDDDKDP 795
QY 1286 KKGKGRP---PKPLGGGP---KEEPTMKTSKSKS-----KSGPARPEE---EEBE 1331
DB 796 NDPYALDIDDKPLADESKLPTQKHRTETSKSPKQDVPWEKSKPKKKEKKHKE 855
QY 1332 ROSGNTQSKSKQHRVSRRAQORAESESSAIESTQSP---QKGRGPRSKTPSPSQ 1386
DB 856 RDKKKKKKKKKSPKPKKKKRRKKEERTKCKKSKKOPPOSEEAGAPVONGADEE 913

RESULT 15
NP14_HUMAN STANDARD: PRT; 699 AA.
AC Q14978; O15030;
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nop140) (Nucleolar and colled-body
DE phosphoprotein 1).
GN NOLC1 OR KIAA0035.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Leukemia;
RC MEDLINE=95386590; PubMed=7657714;
RA Pal C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
RT "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
RT protein p130 are associated with nucleologenesis.";
RL J. Cell Sci. 108:1911-1920(1995).

SEQUENCE OF 3-699 FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid

```

```

RT cell line KG-1.";
RT DNA Res. 1:27-35(1994).
RL [3]
RP ALTERNATIVE SPLICING.
RX MEDLINE=96205319; PubMed=8630004;
RA Pal C.-Y., Yeh N.-H.;
RT "Cell proliferation-dependent expression of two isoforms of the
RT nucleolar phosphoprotein p130.";
RL Biochem. Biophys. Res. Commun. 221:561-567(1996).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=97168979; PubMed=9016786;
RA Chen H.-K., Yeh N.-H.;
RT "The nucleolar phosphoprotein p130 is a GTPase/ATPase with intrinsic
RT property to form large complexes triggered by F- and Mg2+.";
RL Biochem. Biophys. Res. Commun. 230:370-375(1997).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20036810; PubMed=10567578;
RA Chen H.-K., Pal C.-Y., Huang J.-Y., Yeh N.-H.;
RT "Human Nop140, which interacts with RNA polymerase I, implications
RT for RNA gene transcription and nucleolar structural organization.";
RL Mol. Cell. Biol. 19:8536-8546(1999).
CC -1- FUNCTION: RELATED TO NUCLEOGENESIS. MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC -1- TRANSCRIPTION CATALYZED BY RNA POLYMERASE I.
CC -1- SUBUNIT: INTERACTS WITH RNA POLYMERASE I 194 KDA SUBUNIT (RPA194)
CC -1- AND WITH CASEIN KINASE -II.
CC -1- CYTOPLASMIC LOCATION: SHUTTLES BETWEEN THE NUCLEOLUS AND THE
CC PRE-NUCLEOLAR BODIES WHICH ARE SUBSEQUENTLY RELOCATED TO NUCLEOLI
CC AT THE EARLY G1-PHASE.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=Alpha;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha;
CC IsoId=Q14978-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q14978-2; Sequence=VSP_004338;
CC -1- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK2 AND PKC SITES. THERE IS EVIDENCE
CC SUGGESTING THAT CDK2 KINASE PHOSPHORYLATES P130 AT THE M-PHASE.
CC -1- SIMILARITY: Contains 1 Lish domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z34289; CA84063.1; -
CC EMBL; D21262; BAA04803.1; -
CC PIR; I38073; I38073
CC Genew; HGNC:15608; NOLC1.
CC GK; Q14978; -
CC MIM; 602394; -
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005730; C:nucleolus; TAS.
CC GO; GO:0007049; P:cell cycle; TAS.
CC GO; GO:0007067; P:mitosis; TAS.
CC GO; GO:0006364; P:rRNA processing; TAS.
CC InterPro; IPR006594; Lish.
CC Pfam; PF05022; SRP40_C; 1.
CC SMART; SM00667; Lish; 1.
CC PROSITE; PS50896; Lish; 1.
CC Nuclear protein; phosphorylation; Repeat; GTP-binding; ATP-binding;
KW Alternative splicing.
FT DOMAIN 10 42 LISH.
FT DOMAIN 84 566 11 X 12 AA APPROXIMATE REPEATS OF AN
FT ACIDIC SERINE CLUSTER.

```

```

FT REPEAT      84      95      ACIDIC SERINE CLUSTER 1.
FT REPEAT      125     136      ACIDIC SERINE CLUSTER 2.
FT REPEAT      167     178      ACIDIC SERINE CLUSTER 3.
FT REPEAT      221     232      ACIDIC SERINE CLUSTER 4.
FT REPEAT      264     275      ACIDIC SERINE CLUSTER 5.
FT REPEAT      325     336      ACIDIC SERINE CLUSTER 6.
FT REPEAT      363     375      ACIDIC SERINE CLUSTER 7.
FT REPEAT      425     436      ACIDIC SERINE CLUSTER 8.
FT REPEAT      470     481      ACIDIC SERINE CLUSTER 9.
FT REPEAT      519     529      ACIDIC SERINE CLUSTER 10.
FT REPEAT      555     566      ACIDIC SERINE CLUSTER 11.
FT DOMAIN      68      82      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN      204     382      INTERACTS WITH RPA194.
FT DOMAIN      384     587      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN      601     617      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES     563     563      PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT VARSPIC     241     241      K -> KWTITSVRAE (in Isoform Beta).
FT             241     241      /FTid=VSP_004338.
FT CONFLICT     3       3       D -> A (IN REF. 2).
FT CONFLICT    133     133      R -> S (IN REF. 2).
FT CONFLICT    291     292      YA -> SV (IN REF. 2).
FT CONFLICT    456     456      S -> P (IN REF. 2).
SQ SEQUENCE     699 AA; 73720 MM; DED4AD94EDF659FB CRC64;

```

Query Match 2.9%; Score 206.5; DB 1; Length 699;
 Best Local Similarity 22.3%; Pred. No. 0.014;
 Matches 93; Conservative 55; Mismatches 172; Indels 97; Gaps 13;

```

QY 1045 ENIKOTDAOGPPDARKNEKLYCYDYAMNIMSKSTTYSLESPKDYLPARFTQPDKN 1104
DB 228 DDSEEEKRAAATPKKTVKPKQVAKAPVKAATTPTRKSSSSSEDSSEDEEOK--KPMKN 284
QY 1105 FSNTKNTLPPEMKSFTPGKPKTTNVLGAVNKPLSSAGKQSQOTKSRMETVSNASSSN- 1163
DB 285 KPGYSYAP-----PSAPPKKSLG-IQPKKAYEQQPVSS--EDSDSDSDSSE 334
QY 1164 -----PSSPGRIKGRLDSEMDHSEMDYIMSSPLPGKKSDDKDDSD 1205
DB 335 EEKPPTKAVVSKATTKPPAKKAESSSDSDSSE-DDEAPSKPAGTTKNSSNRPVAV 393
QY 1206 LVHSELEKPPGRKKTTPY-----TQEEKLGMDLT-KLYVEQKPKGSORS 1249
DB 394 TTKSPAVKPAAPKQPVGGQKLTTRKADSSSSSEESSSEEEKTKRMVATTKPKATAKA 453
QY 1250 RKRNGTASESDEQWPEEKRLKEDILENEDENSPPKGKGRPPKPLGGGP----- 1302
DB 454 -----ALSLPAKQAPGSGRSDSSSDSSSEEEEEKTSKSAVKKKPKQVAGAAPSKPA 507
QY 1303 -----KEEPTWTKSKGSKKSGPPA-----PEEEE 1328
DB 508 SAKGKAESSNSSSDSDSEEEEEKLKGKSPRPQAPKANGTSALTQNGKAANKSEEEE 567
QY 1329 EEEKOSGNTQKSKSKHRVSRRAQORAESPESSAIE-STQSTPQKGRGRPKTPSP 1384
DB 568 EEKKKAHVVSKSGSLKRRKONEAKAEETPQAKKIKIQTPNTPPKRRKKEKGRASSP 624

```

Search completed: September 24, 2003, 20:17:48
 Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:01:29 ; Search time 31 Seconds

(without alignments)
4315.182 Million cell updates/sec

Title: US-09-512-581b-2

Perfect score: 7193

Sequence: 1 MAHSKRTNDGKITPPGVK.....QKGRGRPSKTPSPQPKKNV 1391

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2698.5	37.5	851	2 T00374	hypothetical prote
2	807	11.2	1579	2 T23142	hypothetical prote
3	632.5	8.8	1506	2 S52957	blind protein - Eme
4	571.5	7.9	1303	2 E96805	hypothetical prote
5	557	7.7	1277	2 S54451	hypothetical prote
6	309.5	4.3	390	2 T43647	hypothetical prote
7	303	4.2	780	2 E96840	hypothetical prote
8	264	3.7	990	2 H86293	protein T24D18.4
9	235.5	3.3	1875	2 S38173	myosin-like protei
10	234.5	3.3	852	2 T06310	hypothetical prote
11	231.5	3.2	2253	2 T30336	nuclear/mitotic ap
12	229	3.2	2663	1 S28261	centromere protein
13	227	3.2	1916	2 S46157	RFL1 protein - yea
14	223.5	3.1	4717	2 T41581	hypothetical colle
15	223	3.1	2139	2 T18296	myosin heavy chain
16	222.5	3.1	1819	2 A71928	cag island protein
17	222.5	3.1	2301	2 T02323	nodulin-like prote
18	221.5	3.1	1927	2 G64585	cag pathogenicity
19	220.5	3.1	1727	2 T50073	myosin-like coiled
20	217.5	3.0	1020	1 QPHU8	neurofilament trip
21	216.5	3.0	1790	2 S67353	transport protein
22	210	2.9	1072	1 A37221	neurofilament trip
23	209.5	2.9	1365	2 T45031	hypothetical prote
24	209	2.9	1242	2 G88480	protein C16A3.7
25	207.5	2.9	854	2 S02003	neurofilament trip
26	207.5	2.9	3488	2 T34418	hypothetical prote
27	206.5	2.9	699	2 T38073	nucleolar phosphop
28	206	2.9	6642	2 T29757	protein UNC-89 - C
29	205	2.8	2094	2 S33124	tptr protein - huma

30	203.5	2.8	2364	2 A56577	microtubule-associ
31	202	2.8	2401	2 T28676	riophy protein -
32	202	2.8	2748	2 S57976	nuclear migration
33	202	2.8	6713	2 B89921	hypothetical prote
34	201.5	2.8	3225	2 I52300	giantin - human
35	200.5	2.8	2269	2 T28677	riophy protein -
36	200	2.8	1837	2 T41023	probable nuclear p
37	199	2.8	1359	2 T34036	hypothetical prote
38	199	2.8	2464	1 Q8MSPI	microtubule-associ
39	199	2.8	3187	2 UC5837	36kD Golgi complex
40	198.5	2.8	1695	2 T19823	hypothetical prote
41	197.5	2.7	1538	2 T29095	cardiac muscle fac
42	197.5	2.7	3259	1 A56539	giantin - human
43	195.5	2.7	1679	2 S48385	hypothetical prote
44	195	2.7	488	2 T46014	cylicin II - bovin
45	194	2.7	374	2 C88734	protein F32E10.6

ALIGNMENTS

RESULT 1					
T00374					
hypothetical protein KIAA0648 - human (fragment)					
C:Species: Homo sapiens (man)					
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000					
C:Accession: T00374					
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998					
A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl					
A:Reference number: Z1142; MUID:98403880; PMID:9734811					
A:Accession: T00374					
A:Status: preliminary; translated from GR/EMBL/DBJ					
A:Molecule type: mRNA					
A:Residues: 1-851 <ISH>					
A:Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110					
A:Experimental source: brain					
C:Genetics:					
A:Note: KIAA0648					
Query Match					
Best Local Similarity 62.1%; Pred. No. 1.1e-121;					
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;					
Qy	477	LYIYATLDNAVAVLDMKCONLHRQKDLDLKOKPTDASVAITSKWVITRNL	536		37.5%; Score 2698.5; DB 2; Length 851;
Db	1	LYIYASLDPAVAVLDMKCONLHRQKDLDLKOKPTDASVAITSKWVITRNL	60		Best Local Similarity 62.1%; Pred. No. 1.1e-121;
Qy	537	PDPGKADPMKKFTQVLEDEDEKIRKOLEVLVSPSCGKQAGCVREITTKLGNKOPTNP	596		Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;
Db	61	PDPGKADPMKKFTQVLEDEDEKIRKOLEVLVSPSCGKQAGCVREITTKLGNKOPTNP	120		
Qy	597	FLEMIKFLERIAVPHIDTESISALIKOVNKSIDGTADDEDEGVPTDOAIRAGLELLKVL	656		
Db	121	FLEMKFLERIAVPHIDTESISALIKOVNKSIDGTADDEDEGVPTDOAIRAGLELLKVL	180		
Qy	657	SFTHPISFHSARETESLACIKMDEKVAEALDIFKNTGSKIEDEPPHRSALLPVLLH	716		
Db	181	SFTHPISFHSARETESLACIKMDEKVAEALDIFKNTGSKIEDEPPHRSALLPVLLH	240		
Qy	717	KSKGPPROAKVYAHCHIAIFSSKEPOFAOIFELHSIDPSNLEHITPLVITGHTALL	776		
Db	241	KAKRGTHQAKOAVHCHIAIFSSKEPOFAOIFELHSIDPSNLEHITPLVITGHTALL	300		
Qy	777	APDQFAAPKSWATFTVKDLMLNDRLLGKRTKILNVPDEVSPTVVKIOAIRMVRL	836		
Db	301	APDQFAAPKSWATFTVKDLMLNDRLLGKRTKILNVPDEVSPTVVKIOAIRMVRL	360		
Qy	837	LGMKNHKSAGTSTLRLLTTLTSHDGLTEQGIKSKDMRSLRLAAGSALVKAIOEPCYH	896		
Db	361	LGMKNHKSAGTSTLRLLTTLTSHDGLTEQGIKSKDMRSLRLAAGSALVKAIOEPCYH	420		
Qy	897	EITLLEQYCALAINDEQYQVQVFAQKHLKGLSRRLPLFEMALCALCAKDPVERRA	956		

Dh 1347 GSSAVGRMKPRKRTKTKTGDDWSSDGEAGAVNASSATRRSRNGS-ASRRI SYADDDSD 1405
Qy 1261 E--QOMPEEKRLKEDILLENEDQNSPPKKGKGRPPKLGCTPEEPTMTKSKGSKK 1318
Dh 1406 EDDMEMELNQRARDEDEGEDQ-----AKDIENGSDLSE-----LSESDSMML 1448
Qy 1319 SGP-----PAPEEEEBEROSGNTQKSKS-----KQHRVSRRAQORAESP 1359
Dh 1449 EEPEDDDQPSKEEDQPDQKONGDDAQPASVPVASKAKVPGKAMKKTLP 1499

RESULT 4

E96805

hypothetical protein T5M16.19 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E96805

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:1130712

A:Accession: E96805

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1303 <STO>

A:Cross-references: GB:AE005173; NID:96382504; PIDN:AE07790.1; GSPDB:GN00141

A:Gene: T5M16.19

A:Map position: 1

Query Match 7.9%; Score 571.5; DB 2; Length 1303;
Best Local Similarity 20.6%; Pred. NO. 1.2e-19;
Matches 294; Conservative 270; Mismatches 566; Indels 297; Gaps 53;

Qy 38 VVKTFMMDDSEEEKETLYN-----LALHLASDFELKHPGKDYRLVAVACLADIER 89
Dh 21 VANTLSKIDOPSAINKEKGLKLIKAEIRPLKSIKHAHLKRNNDVSLTVCSELER 80
Qy 90 IYAEADYTPDKLIDFMETTRQLKGLDTRKSPQFNRYFLLENIAVAKSYNICFEL-- 147
Dh 81 ILAPHLPE-EDEYLRDIFTLFLIAEFSELDTVSPYFSRAKILETVSLK---PCLMLD 136
Qy 148 EDSNEITQLYKRLFSTYINNGH-----NQR-----VHMAMDLMSII 185
Dh 137 EDCODLVHEMFENMFESLVREHHOOLINOKSMKTQORRANTQOTHSLENNILAIMSDVL 196
Qy 186 CEGDTVSOQLDLYLVNPAHKNLNKOAYDLAKALKRTPQALRYPTTFEYFNQVLMGK 245
Dh 197 EE--EANSFVYVILLENLVKGEPTTSGADKLASSLIERCADRIEPLICSTLVTCFMEKD 254
Qy 246 TTSIDSEHVFLLLELYNIDSHLLSVLPQLEFKLSNDNEERLQVAVKLAKKFGAKDS 305
Dh 255 SIGTNDKDSYHEIIFKISLAPQMLAVIKRLQELLTDQDVAKIKALNLAGRIFAQPKH 314
Qy 306 ELAS---QNKPLMOCYIGFRNDIHVPIRLCEVAFASCHLNNHP--DLAKDITEXLKVRS 360
Dh 315 CLSSYVETQDLYAEFRFRESDSKSAEVRMAALCKGCKCYFANPSGNKASGVLFALIOERLL 374
Qy 361 DPEAIIHDVIVSYVTAAKKIDLLVNDHLNFVRERLDKRWKRKEMAGLAIQYKYA 420
Dh 375 DFDDRVTQALVACDIKFMKMYVPLNLSSEASERDKKISYRKALOKLTFEYDYC 434
Qy 421 LQSNAG---KDAKQIAWIKDKLHLHYQNSID--RLVERIFAQYVNVHNETJERM 474
Dh 435 DKSEGDMITTDNEQIQC---KILLICCEKNCEFRSQNLELVLSDDLFPRLPVEREM 491

Qy 475 KCLYLYVATLIDLVNKALENMAMKQNLRHQVKDLIDLKQPKTD-----ASYKAIFSK 528
Dh 492 RHWVQCFAMINHLKLSINSLISQKRRLONELRHCLTMRKAYDNIEAQRKKKSYEVK 551
Qy 529 VMVITRNLPDQKQDFKFKFTQVLEDEKIRKQLEVLVSTPSCQKQAGCVREITKLG 588
Dh 552 ---LSACFPDASEEDLEFKIDRM--RDSIFDVLTLLLE-ELSTNAQIIRKFLKMG 605
Qy 589 NPKQPTNFMIFILERIAPVHIDPTESIALIKOVNKSIDGADDEDEGVPPDQAIRA 648
Dh 606 ---VKHSEFELRLSTKCSPTSFSSHHVCLLNQLCGSTSANT-----QLKAP 651
Qy 649 GLELLKVLSTPHPIPSHSAETFEESILACLKMDKDEKVAEALQIFKNTGSKIEDEPHRS 708
Dh 652 SIKLLVLINMFPSYLRGE--KQFLKLE-ENDSADELIVLSKAPYSVNF--G 704
Qy 709 ALLVLVHLKSKKGGPRQAKVAIHICHAIF--SKETQFAQIPEPLHS-LDSNLEHLTP 766
Dh 705 DYVPLEKVCLEGRSQTKCAVSAISLGSSEKSVSELCMLMSILCGRNIPPTLOS 764
Qy 767 LVITGHIALAPDQFAAPKSWVATFIYKDLMDRLPGKRTKLMPDEVPETMYKI 826
Dh 765 LACVQYSVLEYDN-----IYEDL-----TSYIY-----RYQI 793
Qy 827 QAIKMYRWLLGMKNHNSKSGTSLRLTLTLHSDGD-----LTEGKISKDPMRLR 879
Dh 794 YGLTLYVKSFLPRHQVYVRKIDDLINLIKTKLSQGHGKISMCFLFLEICSEPTGANVR 853
Qy 880 LAAGSAIVKLAQEOCYHEIITLEQYCALINDECYQVRVFNQKLH--KGLSRRLRPL 937
Dh 854 LAAKAAVLLSRKNDLH-----ISPEVFRITLMKSRFYINGF----- 892
Qy 938 EYMAICALCANDPYKERRAHAROCIVKNINVRREYLOKHAANSEKLISLPEYVVPYTIH 997
Dh 893 -----INKATRESRTC-----BDLDGSESLTDS-----PYMTVFLIH 925
Qy 998 LLADHPDYVKQDIEQLKD---VKEC--LMFVLEILMAKEN-----NSHAFIRKAYEN 1046
Dh 926 VLADHPPEPS---EDCRDEHIVARFCGPLFSLVQVLISINNFTIKETAPFLCIPRA 981
Qy 1047 IKOTKDA---QGPDPAK--MNEKLYVCVANNIIMSKSTYSLSBPKDPLPARFT 1099
Dh 982 IKRAEDAVDSKRTNKAKSRIRNALBOSFETIRYIHIFOSQISMDOQCQKDSLA----- 1035
Qy 1100 QPDRNFSNTKYNLPEPMKSFPTPKPKTYNVLGAVNKLPSAGKQKTSRSMET----- 1154
Dh 1036 ---VGESEKVLPLL-----GNQIETSTG-----STEAQNNTRCSRKRTHLGEH 1078
Qy 1155 -----VSNASSSNSSPGRIRKGRIDSSCMD 1180
Dh 1079 ISCNLSLRIVESEIPIKLEHRTTCAKESYKASVSNKITSKHS--GVVASLKDIS--N 1134
Qy 1181 HSE---NEDYTMSSPLDG-----KKSQRDSDSLV---RSELEKPRGRKTYTDBEE 1227
Dh 1135 HGEAIIGORIKLISPTDCFPYGVYERKNSNSHKITIFDNGDVLVCLDSESETLSHE 1194
Qy 1228 KLGMDDLTKLYQEQPKGQSR--RRGHTASEDEQOPEEKRLKEDILENEDQNSPCK 1286
Dh 1195 SMGOGE--RLGKEFESYSGRNCVPEISHTLAKVNAQK--QTTTKQKQKKVPALNPPAA 1250
Qy 1287 KGRGRPPKPLGGGTPEK-----EPTMTSK-----GSKRKS 1319
Dh 1251 KSKKGNDS--GEGSVSEVTDTSNIKSECKLIKSIQAQFGTKRKS 1295

RESULT 5

S54451

hypothetical protein YMR076c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR582.01c; hypothetical protein YMR916.15c

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999

C:Accession: S54451; S52836

R:Gentiles, S.; Bowman, S.

submitted to the EMBL Data Library, May 1995
A:Reference number: S54451
A:Accession: S54451
A:Molecule type: DNA
A:Residues: 1-682 <GEN>
A:Cross-references: EMBL:Z49259; NID:9807956; PIDN:CAA89222.1; PID:9807957; MIPS:YMR076C
A:Experimental source: strain AB972
R:Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, April 1995
A:Reference number: S52814
A:Accession: S52814
A:Molecule type: DNA
A:Residues: 659-1277 <PEA>
A:Cross-references: EMBL:Z48952; NID:9763008; PIDN:CAA88801.1; PID:9763023; MIPS:YMR076C
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:PD55
A:Cross-references: SGD:S0004681; MIPS:YMR076C
A:Map position: 13R

Query Match 7.7% Score 557; DB 2; Length 1277;
Best Local Similarity 20.3% Pred. No. 5.8e-19;
Matches 272; Conservative 267; Mismatches 611; Indels 190; Gaps 46;

QY 12 KITPPGKKEISDK-ISKEENVRLKVVYKTFMDMDODSEEEKEL-VTLNALHLASDFE 68
DB 8 KLRKNSPISTSDOLISTNEILDRKALHEELASLDQDNTDLTGDKTRDA---LVSKRL 64
QY 69 LKHGKDVRLVACCLADIFRIYAPADYSPDKLDMFTROLKLEDTKSPQFNRY 128
DB 65 LKHGVDGIRAFATACCLSDILRLYADADAYTDA-QLTDFIKLVLSQFQGLQDQENGYHIOQ 123
QY 129 FYLLENIAMVSNICELEDSNEFTQLYR-----TLESVYNNHGNQVNHMM 177
DB 124 TYLLTKLEYYSIVLLADLPSSNNLLELFHYDPNKSFPARLFNVIG----- 173
QY 178 VDLASSTICEGDFVSOELDTVL-----VNLVPAHKNLNGQ-AYDAKALKRTRAOAI 229
DB 174 --ILGEVISEDFVPLVRLIFPKFLYNNENIEPEGVNTSDCYEVLISLCTDYSKRM 231
QY 230 EPTTTFPNOVYLM-----GKTSISDSEHVFDLILELYNIDSLLSLVLPQLEKLSN 284
DB 232 SRHLTKYSEIITHATNDNNSRLTYVVKLKLVLRLMETVPELINAVIGFIYHELSSE 291
QY 285 DNEERLQVYKLAMFEGA-KDELSAQNKPMLQCLGTFNFIHVRLECVKFAHCLMN 343
DB 292 NELFRKEATKLIIGILTSYDNLNFVSTHDFKAWISKIADISPDVREWTESTIPQIIAT 351
QY 344 HPDLAKDLTEYLKVRSHDPEAIRHDVYSI---VTAAKKDIL-LVNDHLNLFVRERT 397
DB 352 REDISKELNQLATFTIDSPRVKRTSVMIENKVPVTEIMKNTKATYLSLHARE-- 409
QY 398 LDKRWVRKEMMGIAQIYKKYALQ---SAAGKDAQAIAMIKDKLHIYYQNSIDRLL 454
DB 410 --KKEVELEICINMAKFSNSLNEIERTYONKEIMEIIDIIPSTLYMYIINDLINEQ 467
QY 455 VERIFAQVMPHNLTEFERMKCYLYLVATLDLNAVKALENMKQCNLLRHQYKDDLDL 514
DB 468 VDSITFEITLPEPDNDKRVHRLTLVSHDFKKAFTSPFANARQIKISPAISKYIDSK 527
QY 515 QPKTASVKAIFSKVMV-----ITRNLPDGGKADPFMKKFQVLEDEKIKQLEV 565
DB 528 FLNNGESSSSQGIIVNNKYNOTLOWLASGLSDSTKALDALETIKQF--NDRITYLLNA 585
QY 566 LVSPTCSCQAEGCVREITTKLGNPK-----QPTNPLEMIKELLERIAVH 612
DB 586 CVFNIDIPFLTFKNCYNELVSKIQTPGLFKKYNISTGASIMPRD-LAKYIQIILFRASPII 644
QY 613 IDTSSIALIKQVKNKSIDGTADDEGVPDQATIRAGLE--LLAKVLSSTHPISE-HSMT 669
DB 645 YNVSINISVLLNLSNS-----DAKOLDKRRLLDDISVYNPFLFQDQIRT 689
QY 670 FESLLACL-----KMDKVAEALAIQIFKNTGS-KIEDEPHIRSLPVLNHSKSKGP 722

DB 690 LKTIKDDDDPDAEKNMLSLEELKTYLKASKTLKQDVDFDD--TFEFTKLYDAVBSK 747
QY 723 PROAKYAIHCHAIHAISSKETQFOAI---PEPLHKLSDNSLEHLITPLVYIGHALLAPD 779
DB 748 PEITFYATKLI-ALSPKAEETLKKIKIRILPLDLQDKDYFTSHIILVMEIFKPPHVLND 806
QY 780 QFAAPKSMVATFYIKDMLNDRLPGKKTYYK-WPDEEVSEPMV----KIOAIKMMVR 834
DB 807 D-----STDIIISYLKEVLSNQVGDSSKEELDWEDESLDSTKYSALGNKVFYTKLFTN 861
QY 835 WLGM-----KNHNSGCT-STPLRLITTLILSDGL-TEOGKISKPDMS---RLRLAG 883
DB 862 KLRSLAPVPRDLAESFTEKTKMLPFYLIASGGLISPFKNKEFYPTPSNQTLCRCVAG 921
QY 884 SAIVKLAQPCYHEIITLEQYOLCALAINDECYOYQVPAOKLHGLSLRLPLEYMAIC 943
DB 922 IOVLKARISNINNEIKPSDIIKLINLVEDESLPVKRTLEQLKDYVANELISIKFLPLV 981
QY 944 ALCAKDPYKERRAHARQCLVKNINVRREYIKQHAANVSEKLSLPLEYVVPYTIHLADP 1003
DB 982 FFTAEPVPELTKTTKIMI--NFTGLSFKK-----GTIERALPRLIHAIHNP 1030
QY 1004 DYVKVQDIE-----QLKDVKECLFWLEILMAKNENSHAFIRKVENIK-----QTK 1051
DB 1031 DIVGSLDEBGDAYLALTAIDYLLFYPDSIAQENFSLIYLSRVKNYQDKLVEDLID 1090
QY 1052 DAQGPDAK-----MNEKLYTCDVANITMS---KSTYS-----LESPPDVLPA 1095
DB 1091 EEEGQKDEAKKHNRYQCKMYIIGELSOIMILNKEKKNNQMSAYPGKLMIPSLFKP- 1149
QY 1096 RFTQPDKNFSNTKYLRLPREKSPFTPKPKRTYNLAVNKPPLSAGKOSQSKSRMEV 1155
DB 1150 -FATVQEQOLS-FKTYIPESLH-----EKIONNKAKIGRLHSTQTRQGLKRLAL 1201
QY 1156 SNASSSNPSPSGRIGRISDESMDHSENEYTMSPLPGKKS-D-KRDSDLVRSLEKP 1214
DB 1202 ENNESQKKKKKYNHARQADDEBGDGDRESDDSDSYSPSKNNEFKKGHENTV--MKRL 1258
QY 1215 RGRKTPVTEDEKLGMDL 1234
DB 1259 RVKKEVDYKDDDED--DDI 1274

RESULT 6
743647
hypoetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
R:Lee, M.; Yoo, H.S.; Chung, K.S.
submitted to the EMBL Data Library, February 1998
A:Description: Clone 17 (blnd homologue).
A:Reference number: Z22596
A:Accession: T43647
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-390 <LE>
A:Cross-references: EMBL:AF049529; PIDN:AAD02493.1

Query Match 4.3% Score 309.5; DB 2; Length 390;
Best Local Similarity 24.8% Pred. No. 8.4e-08;
Matches 105; Conservative 76; Mismatches 192; Indels 51; Gaps 11;

QY 825 KIOAIKMMVRMLGKKNHS--KSGTSLRLITTLIHSDDGLTEOGKISKPDMSRLDA 882
DB 4 KVLAIKRVLVNRLRAAGTEALNIGAPIIKLKVLMADGELSPKNTPKISRAVLRLLTA 63
QY 883 GSAIVKLAQPCYHEIITLEQYOLCALAINDECYOYQVPAOKLHGLSLRLPLEYMAI 942
DB 64 SKYFLKLSIPFYAHIDIPSSIVSNLLIMPDENFVNRMLFTKLOKQLOKLLKPLSYPL 123
QY 943 CALCAKDPYKERRAHARQCLVKNINVRREYIKQHAANVSEKLSLPLEYVVPYTIHLAND 1002

```

Db      124 LFTLADPEEEITKA-----SIRSOVAEFQKTHDFTMYVATYLLHLSSH 172
QY      1003 PDYVKYODIEQLDKVCEKLFVLEILMAKNENNSHAFIRKKNVENIKOTKAOQDPDAKM 1062
Db      173 PDLSIESSEMSLDFIAIRFYVDTVNSENV---PIVFLMQRIRKSYDV--IEDG--N 224
QY      1063 EKLYTCQDVAMNIMSKS-----TTY--SLESPKDPVLPAEFYQDPDKNFSTKLYLP 1113
Db      225 NRYIVLSDMAOKILOKVSQNFQMSLTYYPRQIKPIELRPIPSIDKKRIFN--KIFIT 282
QY      1114 PEKSFPTPGKPTTNVLGAVNKPLSAGKOSOTKSRMETVSNASSSSPSPGRKIGR 1173
Db      283 PKHES-----QIEHAIRTPVSSFAKOTTNKHANLKOKKTHSSKSDKSSRRK-- 330
QY      1174 LDSSEMDHSENDYTMSSPLPGKSKDRDSDLVRSLEKPRGKKTPTVEQEKLCMD 1233
Db      331 ---NEKRRKLNQNPNNRNVPERSSSRFGCIRINYSAPS---SSEIESEEEIESED 383
QY      1234 LTKL 1237
Db      384 FDEI 387

```

RESULT 7

hypothetical protein F23A5.16 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 A:Accession: F96840
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: F96840
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-780 <STO>
 A:Cross-references: GB:AE005173; NID:96503292; PIDN:AAF14668.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F23A5.16
 A:Map position: 1

Query Match 4.2%; Score 303; DB 2; Length 780;
 Best Local Similarity 20.4%; Pred. No. 4.3e-07;
 Matches 153; Conservative 118; Mismatches 254; Indels 224; Gaps 27;

```

QY      18 GVAEIDSKISEMVRRLKAVVTFPMDMDSEEEKLYNLAL-----HLASDFLKH 72
Db      18 GTNLSPSPSTDDLTLTLDFTESLKNVED---OPLSMQSAALIPSRNALVSDLLSH 73
QY      73 GKDVRLVAVCLADIFRIYAPAPYTPDKIDFMFTITQGLGLEDTKSPQFNRYFYL 132
Db      74 DSDVRSVSVSCLFEIRITAPERPYSDDLAKKEIFLITIAFEKLADASSRKYKAFVL 132
QY      133 ENIAWVKNICFELEDSENEITQOLYRTFESVINGNHQVHMMDVMSIICEGDTVS 192
Db      133 DINAAYKSCILVMDLE-CYDILLQMFNRNFKFIRSDHPQLVFSMELIMAIIDEPQVS 191
QY      193 QELDLVYVNLVPAHNNKQAVDLAKALKRTAQALEPTITFFPNVOYMLGKTSISDS 252
Db      192 TDLDSLALTVKKENQVSMWSMSLAEKVLSCARKKPYII---EALSKRGTSIDMS 247
QY      253 EHFEDLLELYNIDSHLLSVLPQLEFKLSNDNEERLOY-----VKLAK----- 298
Db      248 PVVSSICQSVFN-----TPKVISPVNTRKEHESLKYVRSESTDAETIGKGRKNSL 298

```

```

QY      299 -----MFGAKDSELASQNKPLWQ-----CYLGRNDIHPIRLCYKFAHCLM 342
Db      299 MNPEDYDISMLSGKRDPLKTNSSNKKIQKSGGVSLGKVPAKKTPLPKE----- 348
QY      343 NHP-DLAKDLTEYK-----VASHPEE----- 364
Db      349 NSPATSSRSILTSLKRSRVKMDSEYDSDLSLSPRLKLASCDFDEPNEDDRKIGNSS 408
QY      365 ---AIRDVIYSIYTAACKDIL--LVND-----HLNPFERFLDRWVRKRAMGLA 413
Db      409 KQPSKNGLEKSQKTAKKVEYAKIYNSSGKRLSANSVAKRRULE---RAPDITLPQS 465
QY      414 QIKRYVALQSA-----GKDAKO---YAWIKDK 439
Db      466 SKRKWVSQVAARLANESEBEPKSHPTRRRYRKEVESDGFEDLVGRVNIWPLDK 525
QY      440 LHIYYONSIDDLRIVERITRQIYVWPNINLETTEMKCLYLYLATLDLNAVKALNEMKQ 499
Db      526 ---TFYEGVID-----SYCTKKMHRVYISDGDSEELNTEERN-- 561
QY      500 NLRHQVKDLDLIKQKTDASVAKIPSKVWVITRNLPDQKADPFKKFTQVLEDEKI 559
Db      562 -----ELLEDTSADEKEL-----DLPEISPLSDIMQ-----QKV 593
QY      560 RKOLEVLS--PTCS-----CKQAGCVREITKRLGNPKOPTNPLEMIKFLERIA 609
Db      594 KSKNVAIVSEPTSSGVRSSRTLMKDKGKRLNKQVETRECKN-----LRSLK 644
QY      610 PVHIDTSSIALIKQVAKSIDGTADDE 638
Db      645 ELNAETDRTA--EEQEVSLAESDDRSE 670

```

RESULT 8

protein T24D18.4 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 A:Accession: H86293
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: H86293
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-990 <STO>
 A:Cross-references: GB:AE005172; NID:96587800; PIDN:AAF18491.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T24D18.4
 A:Map position: 1

Query Match 3.7%; Score 264; DB 2; Length 990;
 Best Local Similarity 17.5%; Pred. No. 4.3e-05;
 Matches 241; Conservative 188; Mismatches 440; Indels 506; Gaps 50;

```

QY      43 MDMDDOSEEEKELYNLALHLASDFLKHGKQVRLVACCLADIFRIYAPAPYSPDK 102
Db      47 VEOULSSVOKALHPRALVAD-LRNPDSDVRSVSVCLTEIMKITAPAPY-NDQ 104
QY      103 LKIDFMETITQGLGLEDTKSPQFNRYFYLLENIAWVKNICFELEDSENEITQOLYTLF 162
Db      105 MKDIFQVITIAFEKLADASSRKYKAVILEYTAKVASSLVMDLE-CDLVYLEMQRFL 163
QY      163 SVINGNHQVHMMDVMSIICEGDTVSQELDLTVLVNLVPAHNNKQAVDLAKALL 222

```

```

Db      164 KIIRDPHQLVLSMETIMITIVIDESSEVPDLELITLTYYKKSDQSDVSPALITVEKVL 223
Qy      223 KRTAOALEPYITTFEYNOVLMLGKTSISDLSSEHVEDLIELYNIDSHLLSVLPOLF--- 279
Db      224 SSCCTCKLQPCIM---EALKSSGTS-----IDMYS---PYVSSICQSEFATY 263
Qy      280 -----KLKSNDEERL---OVY---KLLAKMFGAKDSLSAQNPLMOCYLGREFNDIHP 328
Db      264 QAHNDVPRKDEADEKISEGQVVPDSLEDKLNGLSKRGKRSKASGTRRANGD--- 320
Qy      329 IRLCEVFAASHCLMNHDPDLAKDLEYLETKVRSHDEEALRHDIYISIVTAAKKDILLVNDH 388
Db      321 ---EKVITANGLESSTID-AETASGSTRKRGKMSLNMPEGYSFKTSSSKV----- 370
Qy      389 LINFVREERTLDKRRNRKREAMGLOAYIKYALOSAGAKDAKOIAMIKDLHLHIYONS 448
Db      371 -----QEKELG-----DSSLCKVAAKVP----- 389
Qy      449 IDDLVLVERIFAQYMWPNHLETTBRMKCLYLYLATLIDLVAKALNEMKQCNLHRQYKD 508
Db      390 -----LPSKVQOTNQ-----SVVISLSSSGRAFTGSKRSR- 420
Qy      509 LLDLIKOPKTDASYKAIFSKYMWITRNLPDGAQDFKFFTOVLEDEK--IRKQLEVL 566
Db      421 ---TKMEETHDHSVATQ-----PAKKOT-VKTNPAKEDLTGNSVKKHEDGI 465
Qy      567 VSPFSCQQA--EGCVREITKLGPNKOPNPPLE--MIKLLERIAPHVIDTESIALI 622
Db      466 KTGSSSKKEKADNLAKTSARK-----PLAETMWVKPSGKKL--VHSDAK----- 508
Qy      623 KOVKSIDGTADDEDEGVPTDOAIRAGLELKLVSFTPHISFSAETESLSLACIKMDE 682
Db      509 ---KKNSEGASMD---TRIPQSSKS-----KKKDS 532
Qy      683 KVAEALQIENKNTSGKIEEDPHIRSALEPVLHRSKKKPPROAKYAIHCIAIFSSKET 742
Db      533 RATTPATK-----KSEQAPKSHPKMR-----IAGEV 560
Qy      743 QFAQIFERLHKSIDPSNLEHLITPLVTIGHIALAPDOFAFPKSWATFVYKDLMDR 802
Db      561 E-----SNTNEL-----GEE 570
Qy      803 LPGKRTTKLVNDEEVSPEYVWKIOATKMVWRLLGKNNHKSQSTRLTLITLHSDG 862
Db      571 LVGRVAVWMPDLKKFEGVYKSCRYKKM-----HQ-----VTYSDG 608
Qy      863 DLTEQGIKSPDMSRLLAAGSAIVKLAQEPYHEITLLEQYOLCALAINDECYQVROVF 922
Db      609 DV-BELMLKK--ERFKLIEDKSSASEDKEDDLLESFPLSAF----- 646
Qy      923 AQKHLKGLSRLLLEIYMAICALCAKDPVKERRAHARQCLKVINVRREYIKOHANVSEK 982
Db      647 -----IOREKSKKRIYKIN 661
Qy      983 LLSLLEPVVYVYTIHLAHDPEYVQODIEOLKDYKCECLAFVLEITLAKKNNSHAFIRK 1042
Db      662 ---VER---SSPEVSSMOQTMKKD-----S 682
Qy      1043 MVENIKOTKOAQGPDAKKNNEKLYTVCVANNIIMSKSTYSLESPKDPVLPARFFTOPD 1102
Db      683 VTDSIKOTKRTKGLAKLVNNEP-----ESTTGKLNKSLK-----KNGEPD 723
Qy      1103 KNEFNNTNLYLPEPKSFTTPGKPKTTYNVLGAVNKPPLSAGQSOQTSSRMETVSNASSS 1162
Db      724 KTRGRT-----GKKO-----KVTOAMHRIK----- 745
Qy      1163 NPSSPRIGRLDSESDHSENDYTMSSPLPGKSKDPRDSDLVRSLEKPRGRKTPV 1222
Db      746 -----DQDEDELETKDEBDSKL--GKESDA--EPDMEDHQLPEPHNNVETK 790
Qy      1223 TEOEKLGMDDLTKLVQEQPKGSQSRKRGHTASSEDQOWPEEKLKEDILE--NEDE 1280

```

```

Db      791 TDGEE-----QEAKEPTTAESKTNCEEPNAPETDQGEKHSIKKEPNAEKSDGE 839
Qy      1281 QNSPPK-----GKRGPRPKPLGGTPEKEPTMTKS---KKGSKKSGPPAPEEE 1328
Db      840 EOEAKEPNAELKTDGENOEAKELTAERKATDEEHNKAYDAVEQSKQETVWEPEAGEE 899
Qy      1329 EEEQSGNTQEQSKSKQHRVRRARQOARESESSAIESTQSTPOKGRGRPRPKTS 1383
Db      900 QKSVPEPNAEPKTK-----VEEKESAKBOTADTKLIEKEDMSKTKGEETKETY 949

RESULT 9
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BOL>
A:Cross-references: EMBL:Z28320; NID:q486586; PID:q486587; MIPS:YKR095w
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Teast, 9; 1349-1354, 1993
A>Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:q450550; PIDN:CA51948.1; PID:q450554
A:Experimental source: strain S288C
R:Kobellng, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A>Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549; PMID:8483450
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOB>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C:Genetics:
A:Gene: SGD:MLP1
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Map position: 11R

Query Match      3.3%, Score 235.5; DB 2; Length 1875;
Best local similarity 19.5%; Pred. No. 0.0023;
Matches 301; Conservative 246; Mismatches 602; Indels 391; Gaps 72:

Qy      19 VKEISDKISKE-----EMVRRLKVVYKTFMDQDSEEEKL--YLNLALHLASDFLKH 71
Db      278 IKGLSDSLNSEQKQFSAEMSLKORLVLDLESQLNAVKEELNSIRELNTAKYIADSKQOT 337
Qy      72 PG-----KDVALL---VACCLADIFRI-----YAPEARYTPDKLQTFMTITQOLKLE 118
Db      338 PENEDLKEQLQTEKLAQCEKECECLRSSITDEDEDENENLSAASSDQFIFLKKQOLKER 397
Qy      119 DTK---SPOFNRYFYLLFN--IAWVKSVMICFELDSNEIFQLRYTLFVSYNNGHOKVH 174
Db      398 RTKEHLNQIETFIYELKHPPIINSFK---ERDMLENELENNALALLEHNSNKNKAVK 454
Qy      175 MHAYDLMSITCEBD--TVSOELLDPYLVNLVPAHKNLNKOAYDLAKALLKRTAOAIEPY 232
Db      455 ELNAKNOKLYECENDLQTLTKORLD--LCROIQYLLITNSVNSDSKGLRKEEIQ----- 507
Qy      233 ITTFNOYVLMGKRSI--SDLSHVPLDILLYNI-----DSHLLSVLPQLEKIKSND 285
Db      508 ---FTIONMOEDSDTITSESQKAVTELEVFKNIIQLOEKNAELKLVNRLADKLSKE 564
Qy      286 NEERLQVVKLLAKMFGAKDSLSAQNPLMOCYLGREFNDIHPVIRLCEVFAASHCLMNH 345

```


Db 411 SENEK-NKSVQILPSTSGDETANVSPSPMAEELPEQSPVKPTKANOKKSS----- 461
QY 605 LERLAP-VHIDTESIALIKOVNKSIDGTADDEGEVPTDOAIRAGLELLKLVSTHPIS 663
Db 462 TEEVPPSPASIAITEEVS----- 477
QY 664 FHSATFESLACLAKMDEKVAEALQLEFNKNGSKIEEDPHIRSAALLPVLIHNSKKGPP 723
Db 478 -----EERPTSEPOVTKSGKVA----- 496
QY 724 ROAKYAIHCHAIFFSKETOPAOJFEPLHKSIDPSNLEHLITPLVTIGHIALLPADOPAA 783
Db 497 -----SSSKTK-----PTVP----- 506
QY 784 PWSKVATFIYKDLMLNDLPCKTKTKLWVPDEEVSPEPTMYKTOAIRKMYRMLCKNNH 843
Db 507 -----PSKKSTS-----ETKVAKOSEKKYVG-----SDNA 531
QY 844 SKSGSTLRLLTTLIHSDDGLTEQKISKPMKSRLLAAGSAIVKLAQEPCHETITLQ 903
Db 532 QES-----TKPEEKKKPGRGAI----- 550
QY 904 YOLCALAINDECYOVFAQKLHKSRLPLPYMAICALCANDPYKERRAHARQCLV 963
Db 551 -----DE-----ESLHTSSGDNEKP-----AVSGKLASKSKSEAKQTYE 585
QY 964 KNIN---VRREYIKQHAIVSEKLSLEPEYVVPYITLHLLADPDYVKQDIEQLDVKEC 1020
Db 586 ESPNSNTRKRKSLGGKASGESLVG-----SRKY----- 615
QY 1021 LWFVLEILMAKNENNSHAFIRKMYENIKOTKDAQGPDDAKMNEKLYTCOVAMNTIMS 1080
Db 616 -WMPMD-----QAYYKGVVSY-----DAKKKHLVITDDGQELLYLKN 654
QY 1081 TTYSLESPKPYLPARFTQPDKNFSNTKNYLPPEMKSEFTPGKPTTNVLGAVKNPLSS 1140
Db 655 QKWS----- 662
QY 1141 AGKQSTSSRMETVSNASSNSPSPGRIGRLDSSSEMDHSENDYMSPLPKSKSK 1200
Db 663 S-ELSDDEADQGTQEDBADSTVSGAGSSKAK-----ATPASKSKSK 703
QY 1201 -RDDSDLVSELEPRGRKKTTPYEOEKLGMDDLTLVLOKPK-----GSOSRRKG 1253
Db 704 TSQDDKTASKSKDSKASREBEASSESE-----EEEPKTYGKSSSSSK-- 751
QY 1254 HTASESDQWPEEKRLLEDILENDEQNSPPKRGKRGRRPPKPLGGGTPKEPTMTKSK 1313
Db 752 -----DI-----SSVSKSGSKASSK-----KKEEPSKAT-- 777
QY 1314 GSKKSGP-----PAPEEEEBEEROSGN-TEQSKSKOHVSRRAOQRAESPSSAISTQ 1368
Db 778 SSKSGSGPVKSVPAKSKGKAKSGASASTPASKAKESASESESEETPREPE-----PATK 833
QY 1369 STPOKRGRRPKT 1381
Db 834 AKSGKSGSGSKSKS 846

RESULT 11
T30336
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30336
R:Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A:Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem
A:Reference number: 820828; MIMD:97053784; PMID:8698198
A:Accession: T30336
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2253 <MER>

A:Cross-references: EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g1514671
C:Genetics:
A:Gene: NuMA
Query Match 3.2%; Score 231.5; DB 2; Length 2253;
Best local similarity 19.2%; Pred. No. 0.0046;
Matches 329; Conservative 248; Mismatches 535; Indels 601; Gaps 80;
QY 7 RTNDGKITYPGCVKISDKISKEEYVRKMYVTKFMDQDSE----- 50
Db 272 RENQTELOEPKLEIREK--NESLMIRDTLKOCDMKADKLLERKNDOLABENGEL 329
QY 51 -----EKKELYL-----NTAL 61
Db 330 SKYVDLSNRLAQLDALYETTEBOLSLSMQOKONOLESTLSAAGEKKLEENHIL 389
QY 62 H-----LASDFLKHGKDVRLVACCLADIFRIVAPAYTSPDKLDF-----MT 110
Db 390 QGKISMLMD-QLKEMGEITDMPETGDMGDLKL-----DDLKQELAVLNTQCLSL 438
QY 111 TROLKGLDPTKS-----POPNRYFYLENTAWKSTNICPELDS--NEIFTOYRTL 161
Db 439 KEQIHQMEEEKSTAEVEMEAOKSR-----FESEKGOLOEIVNLQYSL 481
QY 162 FSVI-----NNGHOKYHMHVMDLMSIIEGDTVSOEL-LPTVLV-----N 202
Db 482 SEITFOKERLDNEAQAQEH-----LMQITTLKLEIKLSSVYHKBELKGIH 532
QY 203 LVPARKNLNKQAYDLAKAL--LKRTAQAIE-----PYITTFNOVIMLQKTSISDSE 253
Db 533 KVEERNEKNQLENFMKMGNLIGITQOLESKTEKVDVLRBOQOKLICERDSTLSTINE 592
QY 254 H-----VFDLLELYNIDSLLSVLPOLERKLSNDBERLOVYKLLAKKFG--A 302
Db 593 YKCKRDENSGVLTNKYKTELODHOQTSLVIE--KLASEKELASKQVODDLAKMIGLA 648
QY 303 K-----DSELASONKPLMOCYLGAFNDIHVIRLECVKFAHCLMHNPLDAKDLTYLKV 358
Db 649 KCQINDSENDOSK-----SHAA-----TVESLRAQ 674
QY 359 SHDPEAIR-----HDVYSIVTAAKKDILLVNDHLNLFVRERTLDRKWRVRKAMNG 411
Db 675 LSEQSQSLKYKKVYSNVLVSEENSKLDQLDLSVEESLRHLREH----- 719
QY 412 LAQITKTYALQSAACKDA-----AKQIANIKDKLHIYYONSDIDRLVIRFA 460
Db 720 ---LEKRTKPAASLDLKRISHLEEMKKLSERDEALH---NLDEERTACKTIES 771
QY 461 QYMWPHNLETPRMCLYLYATL--DLNAVVALNE-----MMKC-----ONIL 502
Db 772 QL---KHLE-EFYQANESLQAKLAGSCALIKOREEBEDELISKVYVDINKAKGESOQIA 827
QY 503 RH--OVKDLIDLQPKTDA--SVKAIFSQVAVITRNLPDPGKADPFKKFTQVLED-- 555
Db 828 QNSCHMQGTTELKKTSHDOVQOGLGERSKVLMTAKKASETSQ--LEKINQLGELISA 885
QY 556 -----DEKIR-----KOLEVYVPTSCQAOECVREITTK 586
Db 886 ANACIKREAEKEKLVLSALHSAEIKLIAVOGESERLSHLEFALSNAKODLDC--LAKE 942
QY 587 LGNPQOPNPPELMIKPL---LERIAPVH-----IDTESIALIKOVNS 628
Db 943 LSDEYKKAEEFAMKVLKEONSERIASLESLSKSLAVKRRKCESEKSEVEHLKRO 1002
QY 629 IDGTADDEDEGV-----PTDOAIRA--GLELLVLSFTPIPSHSAET 669
Db 1003 LDDSSQKHEALQAKNIEIKQILINAKKATSDLAIKSEMGAQLQKAVD--THKSEPSALON 1061
QY 670 FES---LLA-----CLKMDEKVAEAL----- 689
Db 1062 ELSRSLDLALKEGEVERLNKEALROEIOQOQOTITLKEEFALAKDKVALQKE 1121
QY 690 ---QIFKNMGSKIEEDPHIRSAALLPVLIHNSK-----GPRROAKYAIHCHIAFFSK- 740

```

Db      1122 IKQOAVATKGA--EKEMAKLKS-----VISEKSKRICLEEDIONOKRDLSCIOEHOOSKL 1175
QY      741 -ETQFQO-----IFERLKH-----SLDSNLEHLTPVLTIG 771
Db      1176 GESQGLALADLEKKCKEOKELICEQAKAEAKTLASEKASVSEBREGIOALEIEIG 1235
QY      772 HIALAPD--QPAAPW-----KSWATPIVKDLMM-DRLPGKKT-----KLMVP 814
Db      1236 KENQKACDLOQKOLSEMAVOEEKTELQALKKELFIKVOLEBOSQTSFTDSGEBALLYLS 1295
QY      815 DEVSPEYVYKICATMMVAVRWLIGMNNHKSSTLTLLTLTHSDG-----LT 865
Db      1296 EAEEROALTEAKEQAEYOKEIEMKNKEVNSLOAEIKLISKVTNEEVSVDPEQRLLK 1355
QY      866 EOGKISKP--MSRLRLAGSAIVKLAEPCYHEIITLQYOLCALAINDECYQVRVFA 923
Db      1356 ETSKSAKLEKKMKOLMELEASFKELLEKCAIDCLTEQONLKGA--DQQRMAVDSLO 1413
QY      924 QKL-HKGLSRLRLPLEYMAICALCANDPYKERARAROCVKNINVRREYLKQAAVSEK 982
Db      1414 OKLSKAETNHTLQOEIOAMQKCAE--KEQICISLO--QNLKSNQSLSEFASLKHS 1467
QY      983 LLSLLEFYVYPIIH---LLAHPDYKVQ-DIEOLK-DYKECLMPEVLETLAKNENSH 1037
Db      1468 YOEILAEERDLMQEHQOEELSHKRLTERFOALEKAKEDMTETV--LLEKELHNOELQIH 1525
QY      1038 AFIRKAVENIKOTKDAQPPDAKM--NEKLYTCDAVAMTIMSKSTYLSLESPKPYLPA 1095
Db      1526 KROSEMSYSTQTSIHQOAVNSOLIGANQSISQISDQAKKLESEMSTLK-EQKHE----- 1579
QY      1096 REFQDPKNEFSNKNYLLPREMKSFPTPGKPTTNVLGAVNKPPLSAGKOSQTSRSMETV 1155
Db      1580 -----EMKTLRL-----QYEKTLREGNKQVQETSLQETV 1609
QY      1156 SNASSSNPSPPRIGKRLDSEMDHSENEDYTMSPLPGKKDKRDSOLVASELEKPR 1215
Db      1610 -----TSKYDH-----VSKYVLKDQ 1624
QY      1216 GRKTPVTEQEEK-----LGMDDLTK-LVQEQPKPSQSR--KRGTASESDQOQPEE 1267
Db      1625 -----KTPQEEKORLLQVQELNKLQSOQEKTIHQOQKLKQREGTHEADK-----SH 1674
QY      1268 KRLKEDLLENEDQNSPPKKGKRGRRPKPLGGSTPKPEEPMKTSKSKSKSGCPPADEE 1327
Db      1675 KRYLE--LESQLEQOQAVEHYKQOMEK-----AKVHYDAKKKKON-----QE 1714
QY      1328 EEEEROSGNTQSKSKSKOHRSVSRARQOARSPPE 1360
Db      1715 LSEELOSHITKQOEHLSKEN-----ADLKASEQ 1742

```

RESULT 12

S28261

centromere protein E - human

N/Alternate names: centromere 312k protein; kinesin-related protein CENP-E

C/Species: Homo sapiens (man)

C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001

C/Accession: S28261

R/Gen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

A>Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A/Reference number: S28261; MUID:93024922; PMID:1406971

A/Molecule type: mRNA

A/Residues: 1-2663 <EN>

C/Genetics: A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865

A:Gene: GDB:CENPE

A:Cross-references: GDB:361164; OMIM:117143

A/Map position: 4q24-4q25

C/Superfamily: centromere protein E; kinesin motor domain homology

C/Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop

F:7-335/Domain: kinesin motor domain homology <KMO>
 F:86-93/Region: nucleotide-binding motif A (P-loop)
 F:486-2183/Domain: coiled coil #status predicted <COI>
 F:92/Binding site: ATP (Lys) #status predicted

Query Match 3.28; Score 229; DB 1; Length 2663;
 Best Local Similarity 18.24; Pred. No. 0.0076;
 Matches 275; Conservative 268; Mismatches 537; Indels 434; Gaps 64;

```

QY      17 PGYKEISDKRSKEEMVRRLKMYVKTPEMDQDSE-----EKETLNLALHASP 66
Db      333 PYNEVS--TDALLKRYR--KEIMDKQLEBVSLETRAQAMEKD--QLAQLEEK 363
QY      67 FFLKHGKDVRLVLVACCLDIFRIYAPAPYTPSPDKLIDF-METROLKGL-EDTKSPQ 124
Db      384 DLLQK-----VQNEKENLTRLVYSSSLTQOELKAKR 417
QY      125 FNRFFULLENIAVAVKSNICFELEDSNEITQLYRILFSYINNGHOKVMMHAYDLMSI 184
Db      418 KRRVYTCLOGRINKKNSN-----YADQFNIPVN-ITTKHKLSINLLREI 461
QY      185 ---ICEGDYVSOELDPTVL-VNLVPAHKNLNKOAYDLAKALKRTAQAIPEYITTFENQV 240
Db      462 DESYCSDSYFSNTLDTLSEITEMNPAIKLLNQENIE----- 497
QY      241 LMLKTSISDLSBEHVDLILELYNIDSLLSVLPQLEFKKSNDEERLQVYKLLAKMF 300
Db      498 -----SELNGLRADYDNLVL-----DYEQLRTEKEEMKELTKBKNDLDEFALERTK-- 545
QY      301 GANDSELASQNKPLMOCYGLRFPNDIHPYRLBECVAFRASHCLMMHPLAKRLTYLKYVRSH 360
Db      546 --KDOEQ-----LHEISMLKVLNVRREYNO 571
QY      361 DPEAIGHDIVISIVTAAKKDILLVNDHLNPFRERTIDRMVVRKRAMGLAQIYKKA 420
Db      572 DLNELL-----SSKELLREKEDQIKKIQETIDSK--LEIKIKDLS-----YS 613
QY      421 LQSAAGKDAKQIAMIKDKLLHITYONSIDRLIVERIFAQYVPHNLLETTERMKCLYUL 480
Db      614 LESIEDPKQKQFLFDEATV-----ALDKARESAFLRESEULEKEKMKELATY 661
QY      481 YATLDLNAVAVALEMKKCOULLRHOYKDDLDLKOPRTDASVVAIFSKVAVIR----- 534
Db      662 TKQME-----NDIQLYOSLEKAKKMOVDLE-----LDSAFELITKLSLDGKV 708
QY      535 -----NLPPGKAQDFMKKFTQVLEDEKIRKQLEVLYVSPSCQKQAEQCVREITKLG 588
Db      709 PKDLCLNLELEKITTDLQKELNKEVEENALBREV-ILTS--ELKSLPSEVRLRKREI- 763
QY      589 NPKOPTNPFLEMIKFLLEIRIAPVHIDTES-IALIKQVNSIDGTADDEDEGVPTQOAT- 646
Db      764 ---QDKSEELHIITSEKDKLSEVYHKSRSVOGLIEIGTKDLDATQSNYKSTDOEQ 820
QY      647 -----RAGLELLKVLVSFTHPISFSAETFEESLACLKMD----- 680
Db      821 NFKTLHMFQKTKYKMLLENEKMAQETIVNSK-----EAKQFSSSLALKTLELSYKT 872
QY      681 ---DEKVAEALQIFKNTGSKIEDPFRHSALLPYLHNS-----K 719
Db      873 QELQELKREVOER--NEMQOLKEQLNRODSPLOQVREKRTLLTERKLOQLEEVKTLQE 920
QY      720 KGPFRQAKVAIHCHALFSSKEFOFOAIFERLYKSLDPSN-----LEHLTPVLTGHA 774
Db      931 KDDLKQLOESLQI-----BRDLKSDIHPTVMNINIDTOBOLRNALLESLOHOETITWL- 983
QY      775 LLAIPDQAAPAAVMSVATFIYKDLMDNRLGKTKTKLWVDEEVSPTVYKIOAKIMVVR 834
Db      984 -----KSKISEVSRNLHM-----EERTGETKDEFO-----Q 1010
QY      835 WLIGMKNHNSKSGSTLRLTLTHSDGDLTEQKISK--PDMSRLRLAAGSAIVKLAQ- 891
Db      1011 KMGIDKKQDLEKKNLT-QTLTADYKQNDIEILQQRKLFSLQEKNEQLQOMLESYIAEKQL 1069

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:20:49 : Search time 236 Seconds

(without alignments)
891.814 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193
Sequence: 1 MAHSKTRTNDCITTPPGVK.....QKGRGRPKTPSPSPKKNV 1391

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters: 566894

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	758.5	10.5	363	US-09-925-302-800	Sequence 800, App
2	231.5	3.2	2835	US-09-885-535-4	Sequence 4, Appl1
3	225	3.1	1743	US-09-882-227-624	Sequence 624, App
4	206	2.9	2243	US-10-118-513A-12	Sequence 12, Appl
5	197.5	2.7	1530	US-10-118-513A-6	Sequence 6, Appl1
6	197.5	2.7	2266	US-10-118-513A-14	Sequence 14, Appl1
7	189	2.6	3907	US-10-171-311-2	Sequence 2, Appl1
8	189	2.6	3925	US-10-171-311-6	Sequence 6, Appl1
9	186.5	2.6	2649	US-10-205-219-169	Sequence 169, App
10	186	2.6	1979	US-10-205-823-419	Sequence 419, App
11	185.5	2.6	2437	US-09-815-242-5834	Sequence 5834, App
12	185.5	2.6	6281	US-09-815-242-12996	Sequence 12996, A
13	185	2.6	780	US-09-770-689A-5	Sequence 5, Appl1
14	184.5	2.6	1814	US-10-128-714-3388	Sequence 3388, App
15	184	2.6	676	US-09-823-187-24	Sequence 24, Appl

16	183.5	2.6	1400	US-09-764-1176-7	Sequence 7, Appl1
17	183	2.5	670	US-09-823-187-86	Sequence 86, Appl1
18	183	2.5	671	US-09-946-374-308	Sequence 308, App
19	183	2.5	671	US-09-823-187-85	Sequence 85, Appl1
20	183	2.5	671	US-10-015-387A-308	Sequence 308, App
21	183	2.5	671	US-10-006-130A-308	Sequence 308, App
22	183	2.5	671	US-10-199-672-346	Sequence 346, App
23	183	2.5	671	US-10-006-112A-308	Sequence 308, App
24	183	2.5	671	US-10-187-749-346	Sequence 346, App
25	183	2.5	671	US-10-194-457-346	Sequence 346, App
26	183	2.5	671	US-10-184-642-346	Sequence 346, App
27	183	2.5	671	US-10-196-747-346	Sequence 346, App
28	183	2.5	671	US-10-015-392A-308	Sequence 308, App
29	183	2.5	671	US-10-017-253A-308	Sequence 308, App
30	183	2.5	671	US-10-173-689-346	Sequence 346, App
31	183	2.5	671	US-10-173-690-346	Sequence 346, App
32	183	2.5	671	US-10-173-691-346	Sequence 346, App
33	183	2.5	671	US-10-173-692-346	Sequence 346, App
34	183	2.5	671	US-10-173-694-346	Sequence 346, App
35	183	2.5	671	US-10-173-698-346	Sequence 346, App
36	183	2.5	671	US-10-173-699-346	Sequence 346, App
37	183	2.5	671	US-10-173-707-346	Sequence 346, App
38	183	2.5	671	US-10-174-569-346	Sequence 346, App
39	183	2.5	671	US-10-174-583-346	Sequence 346, App
40	183	2.5	671	US-10-174-587-346	Sequence 346, App
41	183	2.5	671	US-10-174-589-346	Sequence 346, App
42	183	2.5	671	US-10-174-591-346	Sequence 346, App
43	183	2.5	671	US-10-175-736-346	Sequence 346, App
44	183	2.5	671	US-10-175-742-346	Sequence 346, App
45	183	2.5	671	US-10-175-744-346	Sequence 346, App

ALIGNMENTS

RESULT 1
US-09-925-302-800
Sequence 800, Application US/09925302
Patent No. US2002004941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 800
LENGTH: 363
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (358)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-302-800

Query Match 10.5%; Score 758.5; DB 9; Length 363;
Best Local Similarity 45.0%; Pred. No. 5e-42;
Matches 172; Conservative 56; Mismatches 105; Indels 49; Gaps 11;
QY 965 NINVRREYIKOAAVSEKLSLIPYVYPTHTLHPDDYKVDIDQLDKVRCLEMFV 1024
DB 1 NISTREIKONPAATEKLSLIPYVYPTHTLHPDDYKVDIDQLDKVRCLEMFV 60
QY 1025 LEIIMAKNNNSHAFIRKVENIKOTKDAQGPDKAKMEKLYTVCDVAMNIMSKSTYS 1084
DB 61 LEVLTAKNNNSHAFIRKVENIKOTKDAQGPDKAKMEKLYTVCDVAMNIMSKSALCN 120


```

QY 1186 DVTM-----SSPLPGKSDKRDSDLVRS---ELKPRGRKKTPTVEQEXL----- 1229
| : : : | : : : | : : : | : : : | : : : |
Db 1569 ATVOOQIEMERLNHQAALLLKLTYESDKHTTVODPLMELKLIMDSLEERTLNROHKEGAL 1628
| : : : | : : : | : : : | : : : | : : : |
QY 1230 -----GMDDLTK-----LYOBQKPKGS-----OR 1248
| : : : | : : : | : : : | : : : | : : : |
Db 1629 LALGOFQHALDELLAMLTHTEGLLSEOKRFGVGPRAIEIEIAKHVLYQNDVLAHOSTVFA 1688
| : : : | : : : | : : : | : : : | : : : |
QY 1249 SKRKHHTASED-----EQQWPEERKLKEDILENEDEQNSP-----PKK 1287
| : : : | : : : | : : : | : : : | : : : |
Db 1689 VNRAGNDLISSAGEASNLQNKLEVLNQRW-----QNVLEKTEQRKQOOLDGALROAK 1741
| : : : | : : : | : : : | : : : | : : : |
QY 1288 GKRGK-----PPKPLGK-----GTPKE-----EPTMKP-SKK 1313
| : : : | : : : | : : : | : : : | : : : |
Db 1742 GRHGEIEDIQQMLTTERHLASLPLGGLPETAKEDLVNMEYCAFAFEKETEYSLSMOK 1801
| : : : | : : : | : : : | : : : | : : : |
QY 1314 GSKKSGPPAPEEEROSGNTGQSKSKQHRVSRRAQOARAS 1358
| : : : | : : : | : : : | : : : | : : : |
Db 1802 GQOMLARC-PKSAETNIDODINNLIKEMESVEFKLINERTKLEEA 1845
| : : : | : : : | : : : | : : : | : : : |

RESULT 3
US-09-882-227-624
; Sequence 624, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; FILE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 1743
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 876
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-882-227-624

Query Match 3.1%; Score 225; DB 12; Length 1743;
Best Local Similarity 18.4%; Pred. No. 8.2e-06;
Matches 285; Conservative 217; Mismatches 523; Indels 520; Gaps 71;

```

```

Db 467 -----MKTLLEA-----YNDCLKNAT----- 483
| : : : | : : : | : : : | : : : | : : : |
QY 277 LEFKLSNDNEERLOVYKLLAMPGAKDSLSQNKPLMOCYLGRFNIDHVPRIECCYK 336
| : : : | : : : | : : : | : : : | : : : |
Db 484 -----EERIKCLDIL-----KDELN-----KKSIL-----LNQKVOALDCLKN 518
| : : : | : : : | : : : | : : : | : : : |
QY 337 A-----SHC-LMNP-----DLAKULTEYLAKRSHDPEALIHVIVSYVPA 378
| : : : | : : : | : : : | : : : | : : : |
Db 519 AKTDEERNCLTLNDPEIREKFRRELELOKELOE-----KDKIKNAKTEA 565
| : : : | : : : | : : : | : : : | : : : |
QY 379 KKDILL--VNDHLNFRERTLD-----KRMVRKRAMGLQIYKKYALQ----- 422
| : : : | : : : | : : : | : : : | : : : |
Db 566 EKNKCLGSKSAIEIKQALDCLKNAKTDEERNCLNITQDLOKELLADMSKAYKD 625
| : : : | : : : | : : : | : : : | : : : |
QY 423 --SAGKDAKQIAMIWKDLIHTIYQNSIDRLVERIFAQYVPHNET--TERMKCLY 479
| : : : | : : : | : : : | : : : | : : : |
Db 626 CVSKARNEKEQEC--EKLTPEARKKLEQQL-----DCLKNATDEBRKCLMD 674
| : : : | : : : | : : : | : : : | : : : |
QY 480 LYATL--DLNAVKALN-----EMKCONL-----RHQVDDL 512
| : : : | : : : | : : : | : : : | : : : |
Db 675 LPRDLOSITLAKESLKAYKDCVSOAKTEAKKECEKLTPEARKLLEBEAKSVAYLDC 734
| : : : | : : : | : : : | : : : | : : : |
QY 513 IKOPKIDASVKAIFSKVAVITRNLPDGPKAODPMKKFTQVLED-----DEKIROLEVL 566
| : : : | : : : | : : : | : : : | : : : |
Db 735 VSQAKTEAKKEC-EKLT-----PEAKKLEAKSKVAYLDCVSRARNEKEKECEK 788
| : : : | : : : | : : : | : : : | : : : |
QY 567 VSPSCSQOAGCVBEITKLGKPNKOPTNPLEMIKFLERIPAPHIDETISALIKOVN 626
| : : : | : : : | : : : | : : : | : : : |
Db 789 LTPPEAK-KLEGOALDCLKNAKTDEBRKCLDLPKIDQKV-----LAKESVAYLDCVS 843
| : : : | : : : | : : : | : : : | : : : |
QY 627 KSIDGTADDEGCVLPDOAIRAGLELLAYLSTHPISRSATFESSLACLMDKAE 686
| : : : | : : : | : : : | : : : | : : : |
Db 844 QAKTEAKKECEKLTPEA-RKLEBAK-----KSVAXLDC-----SQ 882
| : : : | : : : | : : : | : : : | : : : |
QY 687 AALQIFKNTGSKIEEDPHIRGALLPVLRHKSCKGPPROAKYAI-----HCIAHIFSKET 742
| : : : | : : : | : : : | : : : | : : : |
Db 883 AKTEAKKECEKLT--PEARKL-----EAKSSVAYKDCVSRARNEKEK 927
| : : : | : : : | : : : | : : : | : : : |
QY 743 QFAQIFELHKSIDPSNLEHLITPVYIGHIALPADQFAABKSWAVATFYKDLMDR 802
| : : : | : : : | : : : | : : : | : : : |
Db 928 K-----ECEKLTPEA-----EAKKLEQ-----VLDCLNK 954
| : : : | : : : | : : : | : : : | : : : |
QY 803 LFGKTKTLKVPD--EVSPEPMVKIOAKMMVRLGKKNHSGSTLRLTLILHS 860
| : : : | : : : | : : : | : : : | : : : |
Db 955 TEADK--KRCVADLPKIDQKVLAK-ESVAYLDCVSRARNEKEK----- 997
| : : : | : : : | : : : | : : : | : : : |
QY 861 DGDILTEQKISKPDMSRLBLAAGSAIVKLAQPCYHEITTEYOYCALAINDCEYQVQ 920
| : : : | : : : | : : : | : : : | : : : |
Db 998 -----ECERLTPEARKKLEBAKESL--KAYRDCLSQARNEERACEKLTPE--ARK 1047
| : : : | : : : | : : : | : : : | : : : |
QY 921 VFAOKLHKL-----SRLRLPLEYMAICALCAKDPVERARHAROCVLR--NINVRE 971
| : : : | : : : | : : : | : : : | : : : |
Db 1048 LLEQEVKYSIKAYLDCVSRARNEKEKECEKLTTEARKFLAKOYLINCEKNGNEERKA 1107
| : : : | : : : | : : : | : : : | : : : |
QY 972 YLKHAAVSEKLLSLPEYVVPYTHLLAHDPDYKVDIEQLDKVKECELEITLAK 1031
| : : : | : : : | : : : | : : : | : : : |
Db 1108 CLKN-----LPKLOQE--NILAK-----ESLAKYKDC-----SQAR 1137
| : : : | : : : | : : : | : : : | : : : |
QY 1033 NENNSHAF-----IRKVENIKOTKDAOGDDAKMNEKLYTVGDVAMNIIMSTIYS 1084
| : : : | : : : | : : : | : : : | : : : |
Db 1138 NEERRACRKLTLPEARKLLE-----QEVKSVAYLDCVSRARNEKEKECEK 1186
| : : : | : : : | : : : | : : : | : : : |
QY 1085 LESPKDPLVPARFTQPKNFEN--TKNYLP-----PEMKSFTTPCKPRTT 1128
| : : : | : : : | : : : | : : : | : : : |
Db 1187 LITPEARKFLAKELQOKKAIKIDCLKNADPNBRAIMKCLDLSDEKIKYIOERAKV 1246
| : : : | : : : | : : : | : : : | : : : |
QY 1129 NVLGAVNKPLSAGOSOTKSRMETSNASSSSPSSBGRITKGLDSE--MDHSEN--E 1185
| : : : | : : : | : : : | : : : | : : : |
Db 1247 ADCLAMARTDEEKRCOMLYSDLIOENKRTONQONLSKTERLHQASSECIDNDDPTD 1306
| : : : | : : : | : : : | : : : | : : : |
QY 1186 DVTMSPLPGKKSDDR-----DDSDLYRSLEKPR-----GRKKTPTVEQEXL 1230
| : : : | : : : | : : : | : : : | : : : |

```

Db 1307 QEAIEQLEGISDESERALLIGIKRQADEVDLIYSDLRNRKTFEDNMAAKGYELPMDEKNG 1366
Qy 1231 MD-----DLTKIVQEQPKGQSRRKRGHTASESD--EQWPEERKLEDILENDE 1280
Db 1367 GDIATINATNDADKISDNI-----VASIEDIAKQYETETKTIDKMLE----- 1412
Qy 1281 QNSPPKGGKRGPRPLGGCTPKPEEPTWTKSKGSKKSGPPADEEEEEEROSGNTQOK 1340
Db 1413 -----AKLAKALGNG--KKDDDEKESKSTAEKAEENKIDKDVAFETAKNISPIA 1460
Qy 1341 SKSKOHVSRRAQORAE--SPSSAIESSTOSTPOK----GRCRPS 1379
Db 1461 LKNKKEKSGEEFVDENGNPIDDKKAEKODETSVPVKAFTGSDPT 1505

RESULT 4
US-10-118-513A-12
; Sequence 12, Application US/10118513A
; Publication No. US20030039935A1
; GENERAL INFORMATION:
; APPLICANT: Taga, Tetsuya
; APPLICANT: Kimura, Naoki
; TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
; FILE REFERENCE: 06501-107US1
; CURRENT APPLICATION NUMBER: US/10/118, 513A
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/JP00/05756
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: JP 11-288738
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 11-288739
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 2000-123721
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2243
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-118-513A-12

Query Match 2.9%, Score 206; DB 15; Length 2243;
Best Local Similarity 18.0%; Pred. No. 0.00021;
Matches 290; Conservative 218; Mismatches 509; Indels 590; Gaps 74;

Qy 195 LLDIVLVNLPVPAHNKLNQAADYDLAKALKRTQAQ-IEBYITTF----- 236
Db 526 LLSRLIDIDOPSSISQEBQLEALISAAIQSSIGLGLTGYIKRWIIIEQOPNSAANLRYLE 585
Qy 237 --FNQVIMLGKTSISDISEHVED--LILELYNIDS-----HLISVLPOL-EFKLSN- 284
Db 586 WTKWKV-VLTKEEDRDLCPPLFDSGCRIFDPTQIOSIQCHLLSNLSTVASCAMEOG 644
Qy 285 -----DNERLOYVKILAMFGAKSELASOKPIMOCYLGRENDIHVPIRECYKA 337
Db 645 IITERGLVDSLKNHVTQLCOY-----AHMVLMFCHSGL-----LREGDDALHL 689
Qy 338 SHCLMNPDLAKDLTEYLKRVSHDPEEARIRHDIIVSIYTAAKKDILLVHDLNLPVREPT 397
Db 690 SRLKYNTPVIONYITTSRQKSERSPRGKMHND-----CLMIDGLVSOLGDE- 735
Qy 398 LDKRMRYRKEAMMG-----LAQIY-----KKYA-----LOSAGK----- 427
Db 736 VEKIMK-RDECGTGRYPASTHALLDIYLDNITEASHALTIYLLDIMSFPKKTDP 794
Qy 428 ----DAKQAMIKDKLI-----HIYQNSIDRLIVERIFAQ-----YVVPNMLE 469
Db 795 IESPTFAISMGOVKLVQGFMLLDHNDYENGLD--LLFHPYAKPASWQSKIIIEAFMS 852
Qy 470 TTEMKKCIYLYL-----ATDDLNAV--KALNEMMKC-QNLLIRQVADL--- 510
Db 853 QGEHQALRYLQTKMPTVSSSNEVYLHLTVLLFNRCVNEAMNLLRONSNRVNIIEELLKHA 912

Qy 511 -----DLIKQKRTDAS--YKAIFSXYVITRNLDPGKAQDPM----- 546
Db 913 YEVCQENGLMEDDLKLFPTNTEQECIAVKPLQSSISV-----ENHEFLVHHQORAN 963
Qy 547 ----KRETOYLED-----DEKIRKQLEVLVSPSCQKQABGVREITTKLGNKPOKPTNP 596
Db 964 YISALKLINOILKNMLMSDRDPRLE-----RSYTR-----NS 995
Qy 597 FLEM-----KFLERLAPVHIDESI-----SALIKQVNSISQIADDED 637
Db 996 ILDOYKILPRVQKLAVERAKPYHLSTSVFHEVSPRPKPLSAPK--KAITGTVLTIRS 1052
Qy 638 EGVV-----TDQAIRAGLELLK-----VLSFTPIPSHASETESLIACL 677
Db 1053 TFIENVLSKIGEVWASHPRNGVSLFNSPKTEQSPVYVSHPRH----- 1096
Qy 678 KMDKQVAEALQI-FKNTGSKTIEDPPHRSALLPVLHKKSKGPRQAKYAIHCIH-- 734
Db 1097 ----ELPEAFVGPISNTSQRI-----SRLDLVHV-----PVQPSQCLEFTQOS 1138
Qy 735 -----AIFSSKETQFOIPEPLHKSLSLPSNLEHLIRPLV--IGHITALLAPDOFAAPWK 786
Db 1139 PTRSPLCLLSSSLPLSSQFKRPHONTSPSELLLEPLLYKAKKSLALSAT-----S 1191
Qy 787 SWAFETIVKDLAMDRLPGKTKTL---WVPDEVSPEWVKLOAIKMYRWMLGKN 842
Db 1192 SGFAEFPPIPSILRS---GPRTPPLASPSLSPGSLNPPRPVKETRISFME---GNMTH 1244
Qy 843 HSKSGT---STLRLLTITLHSDG-----DL-----TE 866
Db 1245 WTRATDTRNTKAVSFSTFHCQGPAAETEMKTSDKNTYPLPDPAPKQKVAESLATH 1304
Qy 867 QGKISKPDMSRLRLAASAIYKLAQCEPCYHEITLEQYOCAL-----A 910
Db 1305 SGRLEKLDVSKEDSTASTRSQISLE--YHDAPSEDLGAVVSPKPASSTELTINST 1362
Qy 911 INDECYOVROYF-----AQLHKLGISRLRLPLE---YMAICALC-- 946
Db 1363 LGTERDNDKAKFKEGTPSPYKQIGTGDAVAEAFSELSRHD--VEBAEASFGVSVCEG 1421
Qy 947 -----AKDPYKERRHAR 959
Db 1422 ETSTNSKTSYLDGIVPIESRTSITADHKESVANTVADVESSGTSKCAPVSESLGQ 1481
Qy 960 QCLKVINVRREYKQAAVSEKLSLLE--YVVP---YTHLHAHPDYKVDIE 1012
Db 1535 E-EAKNLSFPELVPGLAEKLEYNLSTEQFCDDLPDDKQSAECDAAEVDOGELE----- 1586
Qy 1073 MNTIMSKSTYSLESPPDPVLPARFTQDPKDNFNTKANYLPPEKKSFTTGKPKTAVLG 1132
Db 1587 --VAQSNFTLILEEGE-----EAESDSAAANMLPEKSTK----- 1619
Qy 1133 AVNKPILSSAGKOSOTKSRMTEVGNASSSN-----PSSPRIGRLDSSMDHSENEYD 1187
Db 1620 --EKPVCYRPHNDRERTDLPASAVTADQESIKVETLTYVEPVPVVALAENLIDYIKOTRS 1677
Qy 1188 TMSGPLPKKSKKDDSDLVRLSELEKPRGKRKTPTVBOEKL-----GMDDLT----- 1235
Db 1678 KEATPVAAGEAGDEGDGAVIYKAAHSSHLNSTPKTYKPRAEATVNTSQSDQWVSSRLTL 1737
Qy 1236 -----KLQEQKPKGQSRKRGHTASE-----SDEQWPE-----E 1267
Db 1738 RQHALSLANTYSEQEPASAVATPKKRTKIKETPESSERTOSDLKVAPEMOLITAAONPPAPR 1797
Qy 1268 KRLKEDI-----LENDE--QNSPPKGGKGRPKPLGGSTP--KEEPTMTSKGS 1315
Db 1798 RRRKRDVSGGTLTPSSGAVEPEPEPOGTPGRRLRTPQPEAPEETPRTKVLSSVRRGT 1857


```
OY 1316 KKK-----SGPPA-----PDEEEEROSGNTOKSKS 1343
      :|:
Db 1858 PRRLKSVENGOSTIFIELDLCKSEKASHDGTVELRMANLEDTOMMEYKODEHSDQOPL 1917
OY 1344 KOHVRRAQOQRAESPSSAISTO-----STPOKGRGRPSK 1380
      :|:
Db 1918 KRRVREVERSVSVTEEPKLDSSQLPLOTGLDVATPRK-RGRPRK 1963

RESULT 5
US-10-118-513A-6
: Sequence 6, Application US/10118513A
: Publication No. US2003003995A1
: GENERAL INFORMATION:
: APPLICANT: Taga, Tetsuya
: APPLICANT: Kimura, Naoki
: TITLE OF INVENTION: THE Y568 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
: FILE REFERENCE: 06501-107U51
: CURRENT APPLICATION NUMBER: US/10/118,513A
: CURRENT FILING DATE: 2002-04-08
: PRIOR APPLICATION NUMBER: PCT/JP00/05756
: PRIOR FILING DATE: 2000-06-25
: PRIOR APPLICATION NUMBER: JP 11-288738
: PRIOR FILING DATE: 1999-10-08
: PRIOR APPLICATION NUMBER: JP 11-288739
: PRIOR FILING DATE: 1999-10-08
: PRIOR APPLICATION NUMBER: JP 2000-123721
: PRIOR FILING DATE: 2000-04-19
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 6
: LENGTH: 1530
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-118-513A-6

Query Match 2.7% Score 197.5; DB 15; Length 1530;
Best Local Similarity 18.2%; Pred. No. 0.00044;
Matches 293; Conservative 223; Mismatches 603; Indels 493; Gaps 68;

OY 56 YLNLALHLASDFL-----KHPGKDVRLVACCLADIFRIYAPAPYSPDKLNDIF 107
      :|:
Db 14 YPPASLHVLMDVLLDGVTEAKHS-----ITILLDLIMWF-PNKTDPPISEFPYVF 66
OY 108 MFTIRQLKLEDYTSPOFNRYFYLLIENIAWYKSNICELEDSNEITFOQLRTLSVTINN 167
      :|:
Db 67 AISMGOYKLIQG-----FWLIDHNDYESGLDLFHPATAKPLSQHSKIIOAFMSQ 117
OY 168 GHNOKVHMHVDMSSITCEGDPVSOELDTVLY-----NLVPAHKNLNKGAYDLA 218
      :|:
Db 118 GEHQDA-LRTIOTKAPYVSSGNDYILHL--TVLLFNRCQVEMANFLROHCHN----- 165
OY 219 KALLKRTQAILEPYITTFNOVLMIGKTSISDLSEHVEDLILELYNIDSHLLSVLPOL 278
      :|:
Db 166 -----RLNIEELKHMVEQOEMGLMEDLL----- 190
OY 279 FKLSNDNEERLOVYKLLAKMFGAKDSLASONKPRMOCYIGRFNDIIVPRLCEVFKAS 338
      :|:
Db 191 -KLPTTDEQEC-LVKFLQSSASVONHEF-----LVHHLORAN--YVP----- 230
OY 339 HCLMNHPLADKLTLEYLVR-SHDEEAIHRDVIYSIYTAAKDILLVNDHLNFRERT 397
      :|:
Db 231 -----ALKLNQTLKINMNDRDPRLRE-----RSLAKRSI 260
OY 398 LDKRRVRKKEAMGLAQYKRYALOSAGKDAQAQIANIKDKLLHIYVONSIDBLVER 457
      :|:
Db 261 LDQYCKI-----LPVHRRLAITERAKP-----YHLSTSVFRLYSR 296
OY 458 IFAGYVPHNLETTERMKCLYIVATLDLNAVKALNEMWKCONLLR----- 503
      :|:
Db 297 KPPLSAVVKQYVGTGLTRSVFIN-----NVLSTKGEVWASKEPINSTTPENSSKIEEPS 351
OY 504 -----HQVKDLDLIKQPKTSDASVKAIK-----SKVAVI 532
```

```
Db 352 PIYVSLPAPELPEAFGTPISKASOKISRLRLDLVYQVPRPSQOSEFIQOQSSMKSPVLY 411
OY 533 TRNLPDPGKAODFMKKFTQVLEDEKIKKOLEVLVSPSCSKQAEGCYREITKKLGANKQ 592
      :|:
Db 412 SRSILSPSSQLSGSPQAISSRASE-----LHLLTETPLV-VKAKSLAMSVT----- 454
OY 593 PTNPLFKIKFLLEI-----APVHIDTESIALIKOVN-KSIDGTAD 635
      :|:
Db 455 -TSGSEFTPOSILRSTPRSTPLASPPSPGSRSPQRLKETRISFVEEDVHKKWLPAGAD 513
OY 636 EDEGV---PTQOARAGLELLKVLSEFTHPIFSHSAETPESLIACLKMDKDEVAABAQIIF 692
      :|:
Db 514 SKLEVFTHPKCAVAVETEMPKSKDRITTSFPLNSEKHQ-----EMDEGQSLEKLDVS 568
OY 693 K-NTGSKIEEDFPHIRSALLPVLNHSKSGPPROAKYAHCIAHIFSSKEFOAIOIEPL 751
      :|:
Db 569 KGNSSVSTISDETLLEYQDAD-----SPEDLEETVYASKPKSSSTALTINVTBOT 619
OY 752 HKSLDPSNLHLIIP---LVYIGHI-----ALLAPDQA-----APMKSWAVATFIYKD 796
      :|:
Db 620 EKDDKDVFASEVTPPSDLOKOMGNLEDAETKDLLVAALFAFSELNHLSPVQGTASLAPS 679
OY 797 LLMNDRLPGK-KTKRLAWPDEEVSPETVVKIOAIKMWRYMLGKNNHSHKS-----G 847
      :|:
Db 680 V-----YEGKLEFQKSKVP---VLDEGLTSVETTPAIR-----ANDNKMADVLGDG 725
OY 848 TSTLRLLTTLHSDGLTEQGIKSPDNRSRLAAGAIKVLADQ-----PCYHEII 899
      :|:
Db 726 NSSLITISGPPIVSERLNQEVALNLKEDHEVEVGLKESVDLPEEKLPISDSPDPTORIH 785
OY 900 TLEQYQCALAINDECYOVROVFAOKLHGLSRLLPLEYMAI-----CALAKDPVKKRR 955
      :|:
Db 786 VIEQEKLEA---QOSGEARNLSFNEIYPS-GTYLKQYNFPTIDQOFCDLADNKTDAE-- 839
OY 956 AHAROCIVKNIN-----VREYVILKQHA--AVSEKL----- 983
      :|:
Db 840 -----CDLAIEVDGELFVAGNSFTLLBEGEBEVEGDPASSDVLPRKANVTPEKLYVSG 894
OY 984 -----LSLPEYVV---PYTHLALHDPYVVKODIEQKDV----- 1017
      :|:
Db 895 ENDNHQIANTLPPSAVTSQOKSQKVDLPYVEPIKVAIAENMLDVIKDRSKETISDTME 954
OY 1018 ---KECLMFVLEILMAKNENSHAF-----IRKVENIKQTKDA----- 1053
      :|:
Db 955 QSIHETIPLVSQNIHCPTKLVKSAFKTAQETSTJTMNVSQDDVYVSKTRRGORIQNVN 1014
OY 1054 -----OGPDAKMKNEKLYIVCDVAMNIMS-KSTYLSLESPKDPVLA 1095
      :|:
Db 1015 VKSAQOQASADVATPRKMGQSVKRTKRAKELISEASENIYSDVGLPQNOQIPQNSVTPR 1074
OY 1096 RFTQOPDKNFSTKNLYLPEMKSPFTPGK---KTTNVL-GAVNK-----PLSSAGQ 1144
      :|:
Db 1075 RGRRRKEVNOQILENTSSVQELOITGREGSKRLKSSQLLPAVEETTKKVVKSSVYKR 1134
OY 1145 SOTSSR---MEVYS--NASSSNPSPSGITKGRLOSSEMDHSENBEDYTMSPLPGKS 1198
      :|:
Db 1135 TPRRKRSVENQESVEIINDLKVSTVYSPSMIRKLRSTINDASENTG-----NKQD 1186
OY 1199 DKRDSDDLVRSELEKPRGRKTPYUQOE-RLGMDLTLKLVQEQPKGQSRKRGHTAS 1257
      :|:
Db 1187 DKSSDKQURLRKHVRVRRGREVSPDVRSDLSLESQUL--VOAEDMD-SALPRRGRGRK 1243
OY 1258 ESDEQOPEE---KRLKEDILNEDQONSPPKKGKGRPRK-----PLGGG--TP 1302
      :|:
Db 1244 IN-----PSEDVGSKAVVEERSPKKKEAPSIRRRSTRTPAKSENVDVGPALCKSILVP 1298
OY 1303 KEEPPTM-----KTSKSGSKSGGPPAPEEEEROGSGNTEOKSK-----SKOH 1346
      :|:
Db 1299 NEELSNWSSKKKLTKTESOSQKRLSHSVSEERTDETHKETEDEERLLATATSFTRKS 1358
OY 1347 RVSRRAQORA-----ESPSSAISTQOS--TPQKGRGRPSKTPSPSQPKNV 1391
      :|:
```

Db 1359 RSSRTRSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAOLKEIV 1408

RESULT 6

US-10-118-513a-14

Sequence 14, Application US/10118513A

Publication No. US2003003995A1

GENERAL INFORMATION:

APPLICANT: Taka, Tetsuya

APPLICANT: Kimura, Naoki

TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS

FILE REFERENCE: 06501-107051

CURRENT APPLICATION NUMBER: US/10/118, 513A

CURRENT FILING DATE: 2002-04-08

PRIOR APPLICATION NUMBER: PCT/JP00/05756

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: JP 11-288738

PRIOR FILING DATE: 1999-10-08

PRIOR APPLICATION NUMBER: JP 11-288739

PRIOR FILING DATE: 1999-10-08

PRIOR APPLICATION NUMBER: JP 2000-123721

PRIOR FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 2266

TYPE: PRT

ORGANISM: Homo sapiens

US-10-118-513a-14

Query Match 2.7%; Score 197.5; DB 15; Length 2266;
Best Local Similarity 18.2%; Pred. No. 0.00078;
Matches 293; Conservative 223; Mismatches 603; Indels 493; Gaps 68;

QY 56 YLMLAHLASDFL-----KHPGQVRLVLAACCAIDFRYAPAPRTSPDKLDTF 107
DB 750 YPPASHAVLMDYLLDGVTEAARHS-----ITTYLLDITMSF-PKKTDTPIESPTVF 802
QY 108 MFTTRQKLGLEDKSPQFNRYFYLLENIAWVKSYNICFELEDSENIPTQLYRTLFVYNN 167
DB 803 AISMGVQKLIQ-----FWLIDHNDYESGLDILFHATAKAPLSMOSKTIIOAMSQ 853
QY 168 GHNOYVHMWVDMSSITCEGDTVSQELDTVLV-----NLVPAHKNLKQAYDLA 218
DB 854 GEHROA-LRYIQTMKPFVSSGNDVILHL--TVLLFNCEVAEANFLEHOCN----- 901
QY 219 KALKRTQAIEPIYITTFNQVLMIGKTSIDSEHFDLILEYXINDSILLSVLPQLE 278
DB 902 -----RLNTEELKHYEVCQEGMLMEDLL----- 926
QY 279 FKLSNDNEERLQVVKLLAMFGAKDSELASQNKPLMOCYLGRFNDIHPRIEVCVFAS 338
DB 927 -KLPFTDTBEGC-LVNFLOSSASVQNHFF-----LLVHHLQRAH--YVF----- 966
QY 339 HCLMNHDPDLAKDLEVLKVR-SHDPEAIRHDVIVSIVTAKKDIILLVNDHLNFWERT 397
DB 967 -----ALKINQTLKINVMNDRDLRE-----RSIAANSI 996
QY 398 LDKRMVVRKAMMGIAQIYKKYALQSAAGDAKQIAMIWKLLHIYYONSIDRLLVER 457
DB 997 LDQYGI-----LPRVHRKLAIERAKP-----YHLSTSVFRLVSR 1032
QY 458 IFQYVWPHNLETTERKCLYLYATLDLNAVKALNEMKCONLNR----- 503
DB 1033 PKFLSAVPKQVVTGVTLTRSVFIN-----NVLSKIGEVWASKPEJNSTPFPNSKIEEPS 1087
QY 504 -----HOYKDLLDIKQPTDASYKALF-----SKYVVI 532
DB 1088 PIYVSLPAPLPEAFEGTPIISKASQKISRLDLVYGVPRPSCQSEPIQOQSMKSPLYLV 1147
QY 533 TRNLDPBGAKQADMKKFTQVLEDDKIRKQLEVLVYSTCSCQAEGCVREITTKLGNPKQ 592
DB 1148 SRSLPSSSQLKGPSOALSRASE-----LHLETPLV-VKRAKSLAMSVT----- 1190

QY 593 PTPNPLEMIKFLERI-----APVHIDTESALIKQVN-KSIDGTADD 635
DB 1191 -TSGFSEFTPQSLRSTPRSTPLASPSRPGRSQRLKETRISFVEDVHPKPIRPAAD 1249
QY 636 EDEGV--PTDQIRAGLELLKVLSTPHRISFASFPSSLACLMDDEKVAEALQITF 692
DB 1250 SKLEVPFTPKCAVPETEWPKSDRTTFLLSPERKHQ-----EMDEGSQSLKELDVS 1304
QY 693 K-NTGSKIEBDFPHISALLPVLHHSKSGPRPQAKVAIHCIIHAITSSKETQFAQFEEL 751
DB 1305 KGNSSVSTSDETTLLEYQAP-----SPDLEEVFTASKPKSSSTALTITNVTEGT 1355
QY 752 HKSIDPSNLEHTLP--LVITIGHI-----ALLAPDOFA-----APKSMVAPEIVD 796
DB 1356 EKDGDVDVASEVTPSDLOKQMNLEDAETKDLVAAAFSELNHLSPVOGTAEALCAPS 1415
QY 797 LLMNDRLPGK-KTTKILMPDEEVSPEWTWIKQIAIKMVMKLLGKMNHKS-----G 847
DB 1416 V-----YEGKIFQKSKVP--VLDEGLTSVEITYTPAIR-----ANDKSMADVLGDGG 1461
QY 848 TSTRLRLTLTHSDGLTEBGKISKPDMSRLRLAASATVLAQE-----PCVHEIT 899
DB 1462 NSSLTISEGPIVSERRLNQEVALLNKEHVEYGVUKESYDLPEEKLPIISDSEPTQETH 1521
QY 900 TLEVOYLCALAINDECYOVQVFAQRLHKGSLRLRLPLEYMAI--CALCACKDPYKERR 955
DB 1522 VIEQEKLEA--QDSEEARNLISFNELYPS-GTLKIQYNFDITDQFCDLADKKDTAE-- 1575
QY 956 AHARQCLVKNIN-----VREYIKQHA--ANSEKL----- 983
DB 1576 -----CDIAEVDDELVAQSNFTLLLEGEGEVEPGDFASSDVLPRKAAVTAETEKLVCSG 1630
QY 984 -----LSLPEYVY--PYTHLHAPDQVQKODIEQKLDV----- 1017
DB 1631 ENDNHGQIAMPASVAVSDOKSOKVDTLPYPERIKALAEMLDVLKDRSKETITSDTME 1690
QY 1018 ---KECLMFWLEIIMAKNENSHAF-----IRKRVENIKQTKDA----- 1053
DB 1691 QSHERTPLVQSINMCPTRLVKSAFTAQSTMTNMVSVQVDVSSKTRTGORIQNVN 1750
QY 1054 -----QCPDDAKMNEKLYTQVDVAMNIMS-KSTTYSIESKDYVLA 1095
DB 1751 VKSAQOASADVATPRPMQSGSVRRKTKRAKEISEASENITSYDVGLEFQNOOIPONSVTPR 1810
QY 1096 RFTTOPDKNSNKNLPRPMKSFETPGKP-----KTTNVL-GAVNK-----PLSSAGQ 1144
DB 1811 KGRKKKEVNODILENTSSVDEIQITTGRESKRLKSSQLEPRAVEETTKKEVAVSSYTKR 1870
QY 1145 SQTKSSR---METVS--NASSSSNPSPGRIKGRLLDSSEHSENEEDYTMSPLPGKS 1198
DB 1871 TPRIRKRSVENQESVELINDLKVSJTVSPSRMIRKLRSTNLDASENTG-----NKQD 1922
QY 1199 DKRQDSDVANSLEKPRGKKTPTVTEQEF-KLCMDLITKIVQOKRKGSRSKRGHTAS 1257
DB 1923 DKSSQOLRIKIHRRVARGREVSPSDVREDSNLESSOLT--VQAEFDM-SAIPIKRRRPKR 1979
QY 1258 ESDSEQQWPEE---KRLKEDILENDEQNSPPKKGRRPK-----PLGGG--TP 1302
DB 1980 IN-----PSHDVSKAAKKEERSPKKKEADPSIRKRSTRNTPASSENVDGKPALGKSLVP 2034
QY 1303 KEPTPM-----KTSKSGSKKSGPAPAEDEEBEEROSGNTQORSK-----SKOH 1346
DB 2035 NEELSNWMSKKRLTKTESQSKRSLHSVSEERTDMTHKFNEDQERLILASTTKSS 2094
QY 1347 RVSRRAQORA-----ESPSSAIESTQOS-TPQKGRGRPSKTPSPSQPKNV 1391
DB 2095 RSSRTRSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAOLKEIV 2144

RESULT 7

US-10-171-311-2

Sequence 2, Application US/10171311

```

; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatz, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-2

```

QY	564	EVLVSPTSCQAQECVREITKKLGNPKQPI	----	NPFLMIKRLERIPAVIHDTESIS	619		
Db	2602	SAL	----	TLRISLSESOVEMPTSLILEEVOETAEKNVLEKEKRL	2645		
QY	620	ALIKOVNSIGT	----	ADDEDEGVPTDQAIRACLETLLKVLFSNHPISFSAET	--PESL 673		
Db	2646	-----	LOKLISGNEKKQREKKKSPD	-----	VEYLK	-----TTLELPHSNEBSGTFNE 2690	
QY	674	LACLKMDKVA	-EALQIFKNTGSKIEED	-----	PPHRSALPVLHHSKKGPPO	725	
Db	2691	LEALRA	-----	ESVATRAELASYKEKAEKLOEBELLYVETNMTSLQKDSLQVROHLA	-----E 2742		
QY	726	AKYAIHCHALFSS	-KETQAPQIFEPPLHSLDPNSNEHLITPLVYTHIALALAPDOFAA	783			
Db	2743	AKERKSTILEKDETEVEOSKACMEPEPLIKSLASQ	-----	TDGTUKLISSNO	2793		
QY	784	PMKSWATFIKYDLDLMBRLPGKTKTKLMPDEVS	-----	PETWYKIOAKIMMYWML	837		
Db	2794	-----	TPOLIVKNAGIOINLOSESS	-----	EETVELISOFTKIKMO	-ELHAAEIL 2840	
QY	838	GMKNHNSKSGTSLR	-----	LTLLTHLSODLTQOKISKIPDMSRLRLAAGSAIVKLAOERC	894		
Db	2841	DMESRHI	-SETETIKREHYAVOL	-----	LKECGTILKAVIQCRLRSKEGSSIPELAHSDA	2894	
QY	895	Y	-HEITLEQ	-----	YOLCALAINDECYOYVOFAOKLHKLSRLPL	937	
Db	2895	YOTREICSDSGSDMOGILYLTNHSQCPDIASEGSESESATD	SPFKI	-----	KLGLR	-AVHN 2952	
QY	938	EYMAICALC	-----	AKDP	-VKERRAH	-----AROCLYNNIVNRE	Y 972
Db	2953	EGMOVLSITESPYGDEGHSIQOVSFWLEKRYINTTSLKDLITKMOLOREAEYD	3012				
QY	973	LKQHAANS	-----	EKLLSLPEYVVPYTHLHLPDY	-----	YKVOD 1010	
Db	3013	SQSHESPDMRGELLALQOYFLERSVLLAARFETLALGTDTAVGLNCLSEQRIOEG	3072				
QY	1011	IEOLKDYKCELMFVLEILMAKNENSHAFIRKXVENIKOTKDAQGP	-----	DAMKNEK	1064		
Db	3073	VE	-YQAAWCELOKADRSSLSEIOLALHONNGRKITLKRIOESEKPSOLELEVNIQOKOS	3131			
QY	1065	LYTVCDVAMNINMSKSTYSLSLESPKDDVLPARFETPODKNFSNTKNVLP	-----	P	1114		
Db	3132	QMLEMOVELSMKRRATELOEOQLSSEKNVVAELKSE	-----	LAOTKLELTTLKAQIKHLK	3187		
QY	1115	EMKSFTTPGKPKTTNVLGAVNKPLSASGAKOTKSSRMETVSNASSSSNPSFGRIKGRY	1174				
Db	3188	ELEAFRLVEXDKTDEV	-HLNDTLASPOKKSRELQWALE	-----	REKAKL	3231	
QY	1175	D	-SSEMHSNEDEYTHMSPLPGKSKDKRDSLDVRSLEKPRRKRKTPYTEOEKIGMD	1233			
Db	3232	GRSEERKEELEDKFS	-----	LESOKORMLQNLNLEOQKILN	3273		
QY	1234	LTKLVQOKPKPGSORSRKRGHAS	-----	ESDEQWPEEKRLKE	-DILENDEQONSPPK	1286	
Db	3274	QOKIESQRMVLDQALDSEQGRNLEQVLLS	-----	EKRIRIEMSTLDRREELHAQO	3327		
QY	1287	KGR	-----	RGRPKPLGGSTPKPEPTMYKTSKKGKKSOPAP	-----	EEEBEEROSGNT	EOK 1340
Db	3328	SSDGTGOSRPLP	-----	SEDLTKELQOLEEKHSRIVELLNETEYKIDSLQTRQOM	3380		
QY	1341	SKSKO	-HRVSRRAQRAESPSSAISTOS	1369			
Db	3381	EKOKVARKRTLQTEOEANTEGOKKIMIELOS	3410				

QY	453	ILVERIPAQYVNPVHNETTEBMRKCLXYLYATFLDNAYKALNEMKCO-NILRHQVKDLD	511
Db	2501	LL-----QLESTYSAK---DLELTQCCKQIKDMQEQGFETEMIKKRTIVNOK	2545
QY	512	LTIQPKTDASVKAL-----FSKVYVITRNLL-PDPSKADDEMKKFPQVLEDEDERIKKOL	563
Db	2546	IYEEKAAALVYSQULEEVAQVAYAFCCODNQTISEPGRTN-IQNLQMRDEE-LGSDI	2601

APPLICANT: Zhao, Xumei
APPLICANT: Monahan, John
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Gialt, Karen
APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersch, Sebastian
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: OF CERVICAL CANCER
CURRENT APPLICATION NUMBER: US/10/171,311
PRIOR FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 6
LENGTH: 3925
TYPE: PRT
ORGANISM: Homo sapiens
US-10-171-311-6

Query Match 2.6%; Score 189; DB 15; Length 3925;
Best Local Similarity 18.5%; Pred. No. 0.0063;

Matches 283; Conservative 267; Mismatches 564; Indels 416; Gaps 73;

19 VKESDK-----ISKEMVRLKMYVTENDMODSEKEL-----YLIALHLASDF 68
2118 LKEKTDCESELLSKEDLOR-----DIOERNEIEKLEFRVRELEADALLYADTF 2167
69 LKHGKOVRLVACCLADIFRIYAPARYTSPDKLQIFMIFITOLGLGDETKSPQNR 128
2168 QK-----VDRKHFCAVEA---KPELSLEY-----QLDAERPAIDRKEI 2205
129 FYLENIAMWKSYNICPELDSNEITQLYRTLFVINGNOKVNHMYDLMSIIEG 188
2206 TNLSEQLQFRE-----ELEKNNE-----EVQQLHQL----- 2233
189 DYVSQELLDVVLVLAHKNLKNQAVDLAKALIKRAQAIERTTF--FNQVLMIGKT 246
2234 -EIQKKESTRLCELEENKLFKDKMEKLGAIKESDAMSTODQHVLFQFAQIIQKEV 2292
247 SISLSEHVFDDILEYINIDSHLLSVLPQLEFKLSGNDN--BERLQVLLAKMFGA-- 302
2293 EIDQNLQVTKL-----OOOLKITTDNKVIEEKNELIRDLQLETOIECLM 2335
303 KDSLEASQNKPLMOCYLGRENHVPRLCECVKFAASHCLNHPDLAKDLTEYLVKRSHP 362
2336 SDOECVKNRR--EEETIEQLNEVIERLQOELANIGQKTSNMHSLSE----- 2380
363 BEAARH--DYIVSVTAAKKDILLVNDHL--NFVEERTLD-----KRMVYK 406
2381 ADSLKLQDLVVAIEKLTALQOQVETANEMFMKNVLTETNFKNQTLQOELFSLKRESEV 2440
407 EAMMGLAQIKKKYALGSAAGKDAKQAJWKDKLLH-----YYON-----SIDDR 452
2441 EKIOSTIPENSVNVAIDHLSKDKPELEVVLTEDAKLSLENQTYKRSFEENGKSTIINETR 2500
453 LLVERIFQAQVWPHNETTERMKCLLYATLTDLNAVAKALNEMKCO--NLLRHQVKDLD 511
2501 LL-----QLESTVSAR--DLELTQCKYQKQKQOQOQFTEMLQKKTIVMLQK 2545
512 LIRQPKTDASVKAI-----FSQVVMYITRN--PDPEKADDFMKKFTQVLEDEDEKIKQL 563
2546 IVEEKVAAALVSOIQLEAVQVYAKFCODNOTISSEPERTN--IQLNLQREDE--LGSDI 2601
564 EVIVSPTSCQKQAGCVREITKKIKGNPKOPT--NFELMIFELRIAPVNHDTESIS 619
2602 SAL---TLRISELESQVVENHTSILLEKBOVEYIAEKVNLKEKKLLE----- 2645

QY 620 ALIKQVNSIDGT----ADDEDEGVPTDOAIRAGLELLKVLSPTHPISFSAET--FESL 673
DB 2646 ----LQKLEGNKKQREKREKRSPOD-----VEVLK-----TTTELFSHNSDESGFNE 2690
QY 674 LACIKMDEKVA--EALQIQKNGSKIEED-----FPPIRSALPVLHHKSKKGPQQ 725
DB 2691 LEALRA--ESVAARKAELASKEKAEKQOEBLVKETNMTSLQDLQSOVRHLA-----E 2742
QY 726 AKVAIHCHIAIFSS--KETQFAQIFEPDLHSLDPSNLEHITPLVLTGHIALAPDOFAA 783
DB 2743 AKELSLIEKEDETEVOESKACMFEPPLKLSKISASQ-----TDGTLKISSNQ--- 2793
QY 784 PWSWVAFTYKDLMDRLPGKKTTLWVPDEVS-----PETVYKIQAKIMVWML 837
DB 2794 ----TPQILVKNAGIQINIQSECS-----EEVTEIIISQFTEKIKMQ--ELHAAEIL 2840
QY 838 GMRKNSKSGTSTLR--LTTLLHSGDGLTEQCKISKPKMSRLRLAAGSAIYKLAQEP 894
DB 2841 DMESRHI--SETEILKREHYAVQVL-----LKEECGLKAVIQCILRSKSGSIPELAHSDA 2894
QY 895 Y--HEITLBO-----YOLCALINDECYQVQVFAKHLKHSRLPL 937
DB 2895 YQTRICSDSGSDMGOGITLTHSQGFDIASGEGESSESATDSFPKKT--KGLR--AVHN 2952
QY 938 EYMAICALC-----AKDP--VKERRAH-----ARQCLVKNINVRRE-----Y 972
DB 2953 EGMQVSLTESPSYDGEDHSIQGVSEPMLEBRKAYINTISSLMDLTKMQLQDEAEVYDS 3012
QY 973 LKQHAAYS-----EKLSLDEYVVPYTHILLANDPY-----VKVQD 1010
DB 3013 SOSHESFDMRGELLALQOVFLERSVLLAARTELTALGTDAVGLNCLBQRIQOEG 3072
QY 1011 IEOLKDVCECLMFVLEITLMAKNENNSHAFIRKVENIKQTKDQGP-----DAKNNK 1064
DB 3073 VE--YQAMBECLQADRSLSEIQALHAQNMGRKITYLKROESKPSQELLEYNIOOKOS 3131
QY 1065 LYTVCDAVMNIINSKSTYSLESFQDVLPAFRTQDKNFSNTKNYLP-----P 1114
DB 3132 QMLEQVLELSMKDRATELOQLSEKMYVAELKSE--LAQTKLELETTLKAQNHK 3187
QY 1115 EKSFTTPGPKRTTNVGAIVNKPPLSSAGKOSQKSSMELVSNASSNSSPGRIKRL 1174
DB 3188 ELEAFRLVVDKDEV--HLNDTLASQKSRLEQWALE-----KEKAKL 3231
QY 1175 D-SEMDHSENEYTMNSPLPGKSKDRDSDLYRSELEKPRGKKTPTVEQEKLCMD 1233
DB 3232 GRSEERKELEDLKFS-----LESQKQNLQNLMLLEQOKOLNES 3273
QY 1234 LTRLVQOKPKGQSORSKRGHTAS-----ESDEQWPEERLKE--DILENEQNSPPK 1286
DB 3274 QOKIESQRLMYDALQSEQGRNLEQVLLS-----EKVRIEMSTIDRERELHAQLO 3327
QY 1287 KGR--KGRPPKPLGGTTPKEEPTMKTSKSGSKSSPPRP--EEEEEROGNT--EQK 1340
DB 3328 SSGQGSQRPPLP-----SEDLKELQOKLEKHSRIVELLNETEKYKLSIQTRQOM 3380
QY 1341 SKSKQ--HRVSRRAQOARSPESSAIESTOS 1369
DB 3381 EKDRQVHRKTLQTEQEAINTGQKKMHQLOS 3410

RESULT 9
US-10-205-219-169
Sequence 169, Application US/10205219
Publication No. US20030138803A1
GENERAL INFORMATION:
APPLICANT: Warner-Lambert Company
APPLICANT: Lee, Kevin
APPLICANT: Dixon, Alistair
APPLICANT: Brooksbank, Robert
APPLICANT: Pincock, Robert
TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain

```

; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 169
; LENGTH: 2649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Pemphigoid antigen
US-10-205-219-169

Query Match      2.68; Score 186.5; DB 12; Length 2649;
Best Local Similarity 18.8%; Pred. No. 0.0052;
Matches 257; Conservative 232; Mismatches 506; Indels 369; Gaps 69;

QY 193 QELDVLVNLVPAHKNLKNKQAYDLAK-----ALTKPTAQLIEPYIT-----TFNQVLM 242
DB 433 QELAEOLLEENHRA--RLNTEAYRAAMOTOWSWITLQCCVEQHKEKENTAYEEFFMDA-- 488
QY 243 LKRTSISDSEHVFDDILELYNIDSHLSVLEPOLEFKLSNDNEERLQYVVKLLAKMFGA 302
DB 489 ---KEATDYLRNLKDAIQRKYSQDRSSSIHKLEDL-VQESMEKEELLQYKSTIANLMGK 544
QY 303 KDS--ELASQNKPLMOCYIGRFNDIHPRL-----ECKVFASHCLMNP 345
DB 545 AKTIILQKPRNS---DCPL-----KTSIPKAIQDYKQIETIYKDECV-----LANN 591
QY 346 DLAKDLTEKLVKSHDPEAIRHDVIVSYTAKKQILLVNDHLMFVREERTDKMRRA 405
DB 592 HRAK-----WKYISPGNEAMWPSVCFIVPPPKKEAVDLANRIGQYQNVLT--WH-- 641
QY 406 KEAMGLAQIYKKYALQSAAGDAKQAIAMIKDKL--LHYIYONSIDRLIVERIFAQYM 463
DB 642 -FSHIMKSVSWHYLINEIDRIRASNVASIKTLMGEHQVLSNLQSR-----PEDF- 693
QY 464 VPHNLETTERMKCLYLYTLDLNAVKALNEMKCONLRLHOVKDILLDIKOPKTASVK 523
DB 694 ---LEDSQESQ---VFSGSD-----ITOLEKEVNVQKQYQELLSAEREQESYV 739
QY 524 AIF-SKVAVYI-----TRNLDPGKAODFMKPFQVLEDDKIRKQLEVLVSP-- 569
DB 740 NLYISVYRIRLRLENCEDRLIRQIRTPLERDDLHESVRITE-QEKLKELERLDDIG 798
QY 570 --TCSCQ-----AEGCVREITKLGKGNKOPTNPLEMIKFLLEIRIAPHI--DTESTI 618
DB 799 TITNKCEFFFSQAASSVPTLRLSELNVYLQNNNQYSSSTYIDKLTQVNLVAKMTQA 858
QY 619 SALIKOVKSIDGTADDEDEGVPTQDAIRAGLELKVLSFTPIHSHSAETPESLIACJK 678
DB 859 EALVKLYETKL-----CEEAVYADK-----NNIENLISTLK 890
QY 679 MDEKVAEALQIFKNTGSKI-----EEDPHIRSAIIPVLAHNSK----- 719
DB 891 QMRSEVDEKR-QVFAHLEDELQAKAISDEMRKTYERDQDFMWEKADQLVERQNVH 949
QY 720 -----KGPPROAKYAIHCINAI-----FSSKETQAFQIEPLRLKSLDPSMLHLTP 766
DB 950 VOIDNRLRDLLEGIGKSLKYRYRDTYHPLDWIOQVETQRIQGN-----QPEKSKLTAQ 1004
QY 767 L-----VTIGHIALIAP--DQFAAPKMSWATFIVKQLLMDRLPGKKTTKLVMPDEEVS 820
DB 1005 LNOOKLWSEIEIKOSKMDCEQKYAYQYAT--VKDYELQ-----TMYIRAWDSQOKSP 1057
QY 821 EFMVKTQ-AIKMNVRLGLMKNNHSGSTLRLTLTILHASODLLEQOKISKPDMSRLR 879
DB 1058 VRRRRQSSADLIQFMDLRTY---TALVTLMQYIKFAGDSLK--RLSEELIRKCK 1111
QY 880 LAAGSAIVLAQEPCHETI-----TLEQYOLCA-----LAINECVQ 917

```

```

DB 1112 -----FTSEHGAVSDDLQROKATVLENSKLTGKISELERNVAELKKOKSVEBELPK 1163
QY 918 VROVFAOKLHK-----GLSLRLPLEYMAICALCAD-----PKERRAHARQC- 961
DB 1164 VREAENELRKQORNVEDISLOKIRASE-----AKOYRELETIVREKAEARELE 1215
QY 962 LVKNIVVRREYLKQHAASEKILSL--LPEYVVPYTHILMHPDYPVVOIDEOLKDVK 1018
DB 1216 RVNQLTLEAE--AKRAVEENLINFNRQLENT--FTRRTL--EDHLRKDL-SLNDLE 1267
QY 1019 ECLMFLVLEILMAKNENSH-----AFIRKVENIKQTKDAQDPD--DAKME 1063
DB 1268 QOKNKLMELRRKRDNEEELKLKIQEMKDLAQKQVAE--KOLKEKQKITELEARKKITE 1325
QY 1064 KLYTCDVAMNITMSKSTYSLESPPQVLPARFFQPDKNSTNTNYLPPPEKSPFTTG 1123
DB 1326 IOYTCRENAL-----PQCP---IYQATSCRAVWG--HQOE-----HD 1357
QY 1124 KPKTTVGLAVNKLSSAGKOSQTKSSRMETVSNASSNSPSPGRITKGLDSEMDHSE 1183
DB 1358 KQKAEELKQOQVDE-LTAAARKAEDMRRELTIELNALQLEKTS--EKKARLLKDKIDETN 1414
QY 1184 NEDYTSPLPGKSKSDRDSDLVRSLEKPRGRKKTPTVEDEKLGMD--DLTKLVQBO 1241
DB 1415 NTLRLCKLEL-----ERKQDAEKGYSQLRELGRQNLQTTGKAEEAMQASDLKTKRMV 1469
QY 1242 K-----PKGSQSR-----KRGHTASEDDQ-----QMPPEKRL-----KEPI 1274
DB 1470 QLESLNHEKGLQREVDRIYRAHAVAAKNIOHLSQIHSPDEKELRLQIQKRSK 1529
QY 1275 LENEDQNSPPKKGGRPPKPLGCGTPEEPKMTKSKGSKKSPAPAEPEEPEEROS 1334
DB 1530 LKQGFESK-----HEQLQIWK--AEKENNDKIQRLNELEKSN 1566
QY 1335 GNTQ-QSK---SKQHRVSRRAQRAESPESAIESTOSTPQK 1373
DB 1567 ECAEMLKQVYEELTRONNETKLMQRIQASENIVLEKQTIQOR 1610

RESULT 10
US-10-205-823-419
; Sequence 419, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Belia
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatl, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TREATMENT OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0

```

SEQ ID NO 419
 LENGTH: 1979
 TYPE: PR
 ORGANISM: Homo sapiens
 US-10-205-823-419

2.68; Score 186; DB 15; Length 1979;

Best Local Similarity 19.18; Pred. No. 0.0037;

Matches 290; Conservative 226; Mismatches 567; Indels 434; Gaps 65;

Query Match
 16 PGKVEISDKISKEEYWRRLKVVKTMDQDSEEEKLY---LNLALHLASDFLKH 71
 Db PSAAKQSDTMEKRIEILAQSSAYEEVRLQALSDAENIMRLSSLNDNSLAD----- 416
 Qy 72 PGKVVRLVACCLADIRIVAPPEAPYTPSKLKDIFEFITQLKGLDTPKSPQNRFFYL 131
 Db 417 ---NLKLMR-----LEVEKEKSLSQEK-EELQMSLKL-----LNNEEV 454
 Qy 132 LENTAWKSYNICELED-----SNEIFTOYRFLFSYIN--NGHNOKVHHMV 178
 Db 455 IKSTA-TRDISLDELHDLRLNLEAKQDELNQSISE-KETLIAIEIELDRONQATKHM 512
 Qy 179 ---DLMSSICEGDTVSQELDLTVLVNLPRAHKNLRQAVDLARAL----- 221
 Db 513 LIRQLSKQONEGDSIISKLKODLNDEKKRVHQ-LEDDKMDITKELDYQEKLIQSEVAL 571
 Qy 222 --LKRTQATPEYITTFPNQVLMGKTSIS-----DLSEVFDLILELYIDSHLLSV 273
 Db 572 NDHLTKOKLEDKVENLDQLNKSQESVNSIQENLELKEHIRONEEELSRIRRELMOQL 631
 Qy 274 LPQLEFKLSNDNEERLQVVLAKMFGAKDSELASQKPLMQCYLGRFNDIHPIRLEC 333
 Db 632 NQDSNSNPKDOLLKEREAEVRLNKLNL-----SELEQLNENL-----KVAPDYKRN 679
 Qy 334 VKFASHCIMNHDIAKLDTLEYLKYRSHDEPAIRHDIYIVSVTAAKDDILVNDHLLNFV 393
 Db 680 EKLVIAC-----EDVRHQL-----EECLAGNNQL----- 703
 Qy 394 REPTLDKRMVRYKEMMGLAIYKKYVALQSAAGDAKQAIWIKDKILLHIYQNSIDRL 453
 Db 704 ---SLEKNTIVE-----TLMEKGELEA-ELCWAKKRLL-----ESAN 737
 Qy 454 LVERIFAQYVPHNLETTERMKCLYYLATLDLNAVRLNEMKQCNLIRHQVLDLDI 513
 Db 738 KYETIEELSNARNLNLSALQLEHNL---IKLNQKKDM-----ELMELKNNI 782
 Qy 514 KQPTDAS-----VKAFISKVAVITRNLPRPGQAQDPMAKFTQVLEDE 557
 Db 783 EQMDTDHKEKDVLSLSEQKQLTOLINKKEIFIEIKLERSKSLQOELDKYSQALRKNE 842
 Qy 558 KIRQLRVLVSPSCQAEQVREITKLGPNKOPNPFLEMIKFLLE---RIAPYRID 614
 Db 843 ILRQITIE-----EKDRSLGSMKEENNHLQOELERLREEDSKTAPV-AD 884
 Qy 615 TESTISALIKOVNKSIDGTADDEGVPTDQIRAGLELLKVLSTFHPISFASATFESIL 674
 Db 885 PKTLDV-----TELASVSOINTKHELEEKH-- 914
 Qy 675 ACLMDEKVAEALQIFKNTGSKIIEEDPHIRISALLPVLHNKSKGPPROAKVAHCIH 734
 Db 915 -----HOKIIEQONQSKMQLLOSLQEOKKE--MDEFRRQ---H 947
 Qy 735 AIFSKETQAFQIEPRLHKSIDPSNLEHLIRPIVTLIGHIALADQFAAPKSWAVATIV 794
 Db 948 EQMNAHTH---QLF--LEKDEIKSLQKTITQITQLH----- 980
 Qy 795 KDLLMNDRLPGKTKTKLWVPDEEVSPEPMVKAQIKMVRWLLGKN--NHSKSGSTL 851
 Db 981 -----EEHQDIDQNSDIFQET--KVQS-----LNIENSEKHDLSKATE 1019
 Qy 852 RLITTLIHSQD-----LTEQ-GKISKPMDSRLR-----LAAGAIYKLAQEP 893
 Db 1020 RLIVGIERELEIKLNLKNISLTKQIDQLSKDEVGKLTQITLOOKDEIQALHARISSTS 1079

Qy 894 CYHEILTLBOYCALALINDECYQVRFPAOKLHKGLSRRLRLEPYMAICALAKDPVKE 953
 Db 1080 HTQDVVYLOQ-QLOAYMERE-----KVFVNLKERTENSHLKTETHKAMDYAA---AKE 1130
 Qy 954 RRAHARQCLVKNINVRREYLKQNAAYSEKILSLPEYVVPYTHLLANDPQVYODIE- 1012
 Db 1131 AALIKLIDENKMKLSTRESSQD-----MRETIGNLSR---IIRKQIEI 1173
 Qy 1013 -QLKDYECLEMFVLEITMAKEN--NSHAFIRKMWENIKOTDAQSPDDAKNEKLYTV 1068
 Db 1174 DALSOQCOQLTAVLQTSSTGENAGVNSHQF-----ELLQERDKLKQOVKKMEEMQOV 1228
 Qy 1069 CDVAMNIIIMKSTIYSLESPKDPYLPARFPTQDPKNSNKNYLPPEMKSFETPGKPT 1128
 Db 1229 MTTVON--MQHSAQLOEELHQ---LQAVLVSDNNKSLQVQDY--TGLIOSYQNETYKL 1282
 Qy 1129 NVLGAVNRPLSSAQOQOTKS---SRMEFYSNASSSSNPSP-----GRIGRLDSS 1177
 Db 1283 NFGQELAQVQHSIQCLNTDILLGKLDIISPOLSSASILLTPQSAECLRASKSEVLSSESS 1342
 Qy 1178 EMDHSENDYTMSSPLPGKKSDDR-----DSDLVNSELKPPGRKKTPYTQOEK 1228
 Db 1343 ELLQOELBELRKS--LOEKDATTITLQDNHRLSDSIATSELE---RREHETQOSEIK 1396
 Qy 1229 L---GMDLTQVQEO-----KPGQSRKRRGHTASDQQPEE--KRUKEDI----- 1274
 Db 1397 QLKERQVLOKLKLEKDLLIRAKSDOLSSNENNTNKNVNEELLRQAVTYLKEITLLEM 1456
 Qy 1275 ---LENEDQNSPPKGRGRPPKPLGGGTPEKEPTMKTSSKSKRKSPPAPEE--- 1326
 Db 1457 DIGKLKGENEIVETYGKETE-----YQALQETMMKFSMMLREKFECHSKERALA 1509
 Qy 1327 -----EEBERQSGNTED---KSKSKQRRVSRAQQAEE---SPSSAIST--QSTPQK 1373
 Db 1510 FEQLLEKEQOGETELNOLNAVKSQMOEKTVVFOQERDQVVALKQROMENTALQNEVOR 1569
 Qy 1374 GGRPSKTPSPSPQPKN 1390
 Db 1570 LRQKFRSNOBELERLN 1586

RESULT 11
 US-09-815-242-5834
 Sequence 5834, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA-011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR APPLICATION NUMBER: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5834
 ; LENGTH: 2437
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-5834

Query Match 2.6%; Score 185.5; DB 9; Length 2437;
 Best Local Similarity 18.5%; Pred. No. 0.0054;
 Matches 237; Conservative 187; Mismatches 496; Indels 361; Gaps 47;

QY 167 NGHNOKVHMHVMDLSSIIICEDTVS-----QELDTVLNVLVP-AHKRL 210
 DB 1433 NGLNQQ-----OODLHAKKAINNADTVSDVTIDVNNQIDLDAMETLKHLDVNEIPNAEQTV 1488
 QY 211 NKQ-AVLDLAKLL---KRTQAIEPYITTFNQVLMGKTSISDSEHVEFDLLELYND 266
 DB 1489 NYQNDNDNAKTNFDDAKRLA-----NTLNSDNTVNDINGAIGQAVNDATIHNL- 1536
 QY 267 SHLLLSVLPLEFKLSKNDNERLYQVYKLLAMFGAKOSELASQNKPLMOCTLYGRENDIH 326
 DB 1537 -----NGDQRLQ-----DAKRAIQSIN---QALANKLKEIE 1565
 QY 327 VPIRECVFASHCLMNHDPDLAKDLLEYLVKSHDEEAIIRDVIVSYTA---AKKDI 382
 DB 1566 -----ASNATDQDLTKAKNAEELANSITNNIKATSNQAESQV 1604
 QY 383 LLVNDHLNFVREERTLDKRRVYRKEMGLAQIYKKYALQSAAGDAKQAIQIMIKDKLH 442
 DB 1605 QTAGNHAIQVHANEIPK-----AKI-----DANKVDQVQVALIDEIR 1644
 QY 443 IYQNSIDRLVERIFAOYMPHILETTERMKCLYTYATLDLNAVVALNEMKQCNLL 502
 DB 1645 NPNLTDEKQALKDRINQILOQGHN-----GINNATKEEIEQAKOL 1687
 QY 503 RHQVLDLIDLIK-----OPKTDASYKAIEKVMVITRNLPDQKQODEKFKTQYV- 553
 DB 1688 AQAIDPIDLVAKEDAKODVQKQVALIDETIDQ-NPNLTDEK-QALKDRINQILOQGH 1745
 QY 554 -----EDDEKIRKQLEVLVSPSCSQQAEGCVRREITKKLGPNKOPTNPLEMIKPL 604
 DB 1746 NDINNALFKEEIEQAQALQALQIDIKDLVAKEDAKNAIKALANAKR----- 1793
 QY 605 LERIRVHIIDTESISALIKQVNSIDGTADDEGEV-PTDQALRAGLELLKVLSTPHIS 663
 DB 1794 -----DQINSNPDLPPEQAKAKLEIDEAKRALQ-----N 1824
 QY 664 FHSAPFESLACLKMDDEKVAEALQIFKNTGSKIIEEPPIIRSLRLPYLHKKSKKGP 723
 DB 1825 VENAOTIDQLNGLNL-----GIDDIRNTHWEVDQP----- 1857
 QY 724 RQAKAIIICHAIFESKETOFAOIFEP-L-HKSLDPSNLEHLTPYITGHIALAPDOFA 782
 DB 1858 -----AVNEIFEPTEQILLVNGELIYH-----DITTEODIHAHMLT--DQIS 1900
 QY 783 APKSWVAFFIVKIDLLM---DRLPGKKTTLWVPDEEVSE-TMVKIQAIKMMVRYML 837
 DB 1901 AEVIDTPSTATISDILAKVEVTLIDGSKI-VNVPVKVKEKELSVKQQALESJENNAQ 1959
 QY 838 GKKNNHKSQSTLALLTTILHSDGDLPEQGISKPDMSRLAAGSAIVKAQEPCHYE 897
 DB 1960 OKINEINNSVYTLT-----EQKEAIALEVYKLT---QQAIDHNDAPDIVS 2002
 QY 898 IITLEOYQCAL-AINDECYQVQVFAOKLHGLSRLRLPLEYMAICALCARKDPYKERRA 956
 DB 2003 VEIIQOQOAHIEQGNPPOFTIEQ-----AKSNMISTE 2036
 QY 957 HAROCLVKNINVRREYLKQHAANVSEKLSLPEYVVPYTIHLLADPDYVQVODIEQLKD 1016
 DB 2037 DAIOHMEIDIKARTDLTDEKQEAIAKINQLEKQAIQ-----AIQRAQSIDE 2083

QY 1017 VKECLEFVLEIILMAKNENNSHAFIRKVENIKOTKDAQGPDADKMKNEKLYTCVAMNII 1076
 DB 2084 ISEQLQFKAQKKAANPTAKELAKRKQ-EAISRIDFS-----NEKINSTRNSEIGTA 2135
 QY 1077 MRSKSTYSLESPPDYLPAFFTOPDKNFSNTRKYLPPBMSFFTPGKRTTNVLAQVNR 1136
 DB 2136 DEKQA-AMNQINEIYLE-----TIRDINNAHTLQ-----QVEAALNNGIAR-----ISAVQI 2181
 QY 1137 PLSSAGKOSQTKSSRRRETVSNASSSSNPPSRIGKGRDLSSEMDHSEMDYMWSSPLPK 1196
 DB 2182 VYSDAKOS-----SSYGNESNSHLTIGYGTANHPFNSYTIQHK-----K 2221
 QY 1197 KSDKRDDSDLVSELEKPRGRKTPYTBQEEKIGMDDLKLYOQEKPKGOSRRRGHTA 1256
 DB 2222 KUDDEDIDPLHM-----RHFSNNFGNVYKNAIGVAGISGLASFMEFTAKRRK----- 2271
 QY 1257 SESDEQWPEERKLEKEDILENEDEONSPPKRGKRGPRPLGQGTPEKPEPTKTSKGSK 1316
 DB 2272 -EDEEBELIIRUNNKDSIKETLDDTRKHLPLFVKRRKDEEDDYVEEKDSLNGESLDK 2330
 QY 1317 KSGP-----PAPEEEEEEQSGNTEOK-SKSKQHRVSRQAQRAESPESALSTOS 1369
 DB 2331 VHTPEFLPKRRRKDEDEVEYTNENTDEKVLKDNHSPILPAKRRKKEED--VEYTT 2388
 QY 1370 TPQKGRGPRSKTPSPQPKN 1390
 DB 2389 IESKDEDVPLLAKKKKQKDN 2409

RESULT 12

US-09-815-242-12996
 ; Sequence 12996, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12996
 ; LENGTH: 6281
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-12996

Query Match 2.6%; Score 185.5; DB 9; Length 6281;
 Best Local Similarity 18.5%; Pred. No. 0.021;
 Matches 237; Conservative 187; Mismatches 496; Indels 361; Gaps 47;

QY	167	NGHOKVAMHNV	DLMS	STCEG	DVYS-----	QELDPVYLVNV-ARKNL	210					
Db	5275	NGLMQO-----	QODLAHKA	INNADVP	SDVTIDVNNQID	NDAMEFLKLHLYNEI:PNARQTV	5330					
QY	211	NKO-AYDLAKML---	KRTAOA	IBPYITTF	PNQVIML	GKTSISDLSHVPDILJEL	266					
Db	5331	NYQNAADNAK	INFPD	AKRLLA-----	NTLNSD	NVNDINGAIOAVDAIHNL-	5378					
QY	267	SHLLSVLP	OLEFKLS	NDNEER	LOVYKLLAKME	GADELSAONKPLMOCYIGRF	326					
Db	5379	-----	NGDQRLO-----	DAKDA	KAIOSIN-----	QALANKLEIE	5407					
QY	327	VPIR	ECYKFS	HCLMHNP	DLAKULTE	LKRSHPDEALRHDYISVIA	382					
Db	5408	-----	ASATDODK	LIANKKE	ELANSIINN	INNTSNOASOV	5446					
QY	383	LLVNDHLL	NF	REERTLD	KRMVRVR	KEAMGLOIYKKA	LOSAAKDAKOIAWIKDILH	442				
Db	5447	QTAGHNA	LEQV	HANEIRK-----	AKI-----	DANKDV	KOVALIDELDR	5486				
QY	443	IYONSID	RL	VERIFAQV	WVPHN	LETTERMKLYLVAT	LDLNAVAKALNEMKCONL	502				
Db	5487	NPNL	LDKEQAK	LDKRINO	ILQO	HN-----	GNNM	MTKEIEQAKOAL	5529			
QY	503	RHOYK	DL	LDLTK-----	QPKTDA	SYKALF	SVWVITRNLPRPGAQDP	MKFTYOVL---	553			
Db	5530	AQALOD	IT	KVAKEDAK	ODVYDKOVAL	IDIBDIO-NP	RLDTER-KQAL	KRINOILQO	5587			
QY	554	-----	EDDEK	IRKOLEV	YASPTSC	QACG	SVREITFKLNP	POPTNPELMIKFL	604			
Db	5588	NDINN	AL	KEIEQAKOAL	ODINDV	YAKEDAK	NAIKALANAKR-----		5635			
QY	605	LERIA	PV	HDIES	ISALIK	OVYKSID	GTADDEGV-PTDO	IRAGLELLKVL	SPTHIS	663		
Db	5636	-----	DQINS	NPDLTP	PEQAKALKEI	DEAKRALO-----	N		5666			
QY	664	FHSA	TF	PSLLAC	KMDKEKA	BAEALQIF	KNTGSKIED	PHITSALLPVL	HHSSKGP	723		
Db	5667	VENA	OT	IDLQ	NRGLNL-----	GLDID	IRNTHWE	VEDEOP-----		5699		
QY	724	ROAK	KA	IHCHIA	IFSS	KETOPAQIF	EEPJ-HK	SLDPSN	LEHITPLVTIG	HALLAP	782	
Db	5700	-----	ANVEI	FEAT	PEQILV	NGELIVHR-----	DDIT	QODILAN	HLNLI--DOLS	5742		
QY	783	APWK	SWAT	ETV	KDLMN---	DRLP	GKRTTKLV	WPDEE	SPE-TWY	KOAIKMW	837	
Db	5743	AEVID	PT	STATIS	DSLTAK	VEVTL	LDGSKI-VNV	PVYKE	LSVYKQO	IESIENAAQ	5801	
QY	838	GKNN	HS	SG	STLR	LTLTII	HSODLT	EOGKIK	PDMS	SRILAGS	AIYKLA	897
Db	5802	QKINE	INNS	VLTTL-----	EOKEA	IAEY	NKLK---QOA	IVHND	APVHS		5844	
QY	898	IIT	LE	YOYOL	AL-AIN	DECYO	VOVPO	FAOKL	HGSLRL	PLEV	MAICAL	956
Db	5845	VEEL	IOQ	QDAHIE	QNP	EQFTIEQ-----						5878
QY	957	HAROC	LV	KNINVR	REY	LKQHA	AVSEK	LSL	PEYV	VPYTIHL	LHND	956
Db	5879	DAIQ	H	IMDEI	KAR	FDL	DP	KEQO	AI	AKNLQ	KEQAOIQ-----	5925
QY	1017	VKEC	IM	VE	LE	LILAK	NENNS	HA	IFRK	YENIKOT	KDAOG	956
Db	5926	ISEOL	EP	OKAOM	KAN	PTAK	ELAR	KO-EAIS	RIK	DES-----	NEKINS	5977
QY	1077	MSK	ST	TVS	LES	EP	KD	PVLP	AR	FFQ	PD	1136
Db	5978	DEKOA	-----	AMNO	IN	ELVE-----	TIRD	INNAT	ILD-----	QVEA	ALNNGAR---	6023
QY	1137	PLSSA	G	OSQ	OS	TKSS	SMET	VSNASS	NP	SPGR	IKGR	1196
Db	6024	VTS	D	RAKOS	-----	STG	EN	SH	LIT	IG	CTAN	6063
QY	1197	KSDR	K	DDSD	DL	V	SELE	KPR	GR	KKT	PV	1256

```

Db      6064 KLDEDDIDPLMH-----RHFSNNGVNIKALGVGIGSLASFWFIARRRK----- 6113
Qy      1257 SESDQWPEERKLKEDILLENEDQNSPPKGGKRGPRPKPLGGGTPKKEEPMKTSKGGK 1316
Db      6114 -EDEEELEINDNNKDSIKETLDTDKHLPLFLVKKRRKDEDEDYVEEKDSLNGESLKD 6172
Qy      1317 KKSGF-----PAPEEEERQSGNTEBQK-SKSKOHVSRRAOQRAESPSSAISTOS 1369
Db      6173 VKHPEFLPKRRRRKDEDEDVEVTNENDEKYLKDNEHSPLFLFAKKRKDEED--VETTS 6230
Qy      1370 TPQKGRGRPKTPSPSPQPKN 1390
Db      6231 IESKEDVPLLLAKKKKKDN 6251

RESULT 13
US-09-770-689A-5
; Sequence 5, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: C1001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 780
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-770-689A-5

Query Match      2.6%; Score 185; DB 10; Length 780;
Best Local Similarity 18.8%; Pred. No. 0.0011;
Matches 154; Conservative 131; Mismatches 297; Indels 236; Gaps 37;

Qy      682 EKVAEALQIFKNTGSKIEEDPPHRSALLPVLHHKSKGPPROAKVAIHCI---HAIFS 738
Db      23 EYLSDDLQ-----IERLDIVRS---MCHHSKR-----LIACFGOGHTDA 62
Qy      739 SKE-----TQFAQIFERLHKSIDPSNLHLLTPVLTIGHIALAPDQFAAPKWSVATF 792
Db      63 FRRHKPLPLTLAQMWAQSAQLBESLGLKM---LETGCD---AENGLARE-LSQHEVF 114
Qy      793 IYKDLIMNDRLRGKTTKLTWVPDEVSSEFYNKIKQIALIMMYRV--LLGKNNNSKSTS 849
Db      115 VEKEIM--DPLYG-----IAEVEIPINQOKOKOLARVLVDWDSVRAWMQAQHSSTGN 165
Qy      850 TLRLLTTLHSDGDLTEQGKISKRPMSRYRLA-----GSAIVKL--AQEICY 895
Db      166 FQGLRSKIDTLKEENDEAG--NKVEQCKDQLAADMYNMAKEGEYGRFFVTLLEQADYH 223
Qy      896 HEITLLEYOICALAINDEYQVORVFAQKLHKGLSR---LRPLLEYMATICALCAKDPV 951
Db      224 RKALAVLEKALREMAHQDKMAEKPAFTPLSEHLKRSRGRELALPIE--ACVMLLLETGM 281
Qy      952 KER--RAHAROCIVKNNINREYVKQH-----AAVSEKLSLDEYVVVPIYH 997
Db      282 KEEGLFRIGAGASKALTKLAALDCSTSHLDEYSDPAHVAALAKLYLRELPRLMTFSLY 341
Qy      998 LLAHPDVPVKQDIEQLDKVCELMFVLEILMAKKNENSHAFIRKMNENIQTKDAQPD 1051
Db      342 -----EEWTOVASVODODKLYQIYMTTQCKLRPPQNFV-FRLLIKFLAKLQTSVNV-- 392
Qy      1058 DAKNMEKLYTYCDVAM---NIIMSK-----STTYSLESKPDVLP-ARFETQPD 1102
Db      393 --KM-----TFSNIAIVLGPNIILMAKQEGSTLAEIAAATSVHVAVEPIIIOHADMFPEGE 445
Qy      1103 KNFSSTKNVLP-----PEMSSTFTPGPKR-- 1126

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2003, 20:16:09 ; Search time 25 Seconds

(without alignments)
2354.176 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRNDGKITTPGCVK.....OKGRGSPKTPSPQPKKNV 1391

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCITUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	232.5	3.2	2662	4	US-09-595-684B-31
2	201.5	2.8	3248	1	US-08-353-700-1
3	201.5	2.8	3248	5	PCT-US95-16216-1
4	194.5	2.7	2482	1	US-08-328-254-6
5	190	2.6	10182	4	US-09-134-001C-3159
6	185	2.6	8991	4	US-08-714-741-32
7	181	2.5	1780	1	US-08-769-309A-5
8	181	2.5	1780	3	US-08-994-570-5
9	178.5	2.5	2368	1	US-08-198-446B-15
10	178.5	2.5	2368	2	US-08-870-693-15
11	176	2.4	1319	2	US-08-290-731C-2
12	176	2.4	1336	2	US-08-290-731C-6
13	174	2.4	1786	3	US-08-973-462-8
14	172.5	2.4	1865	1	US-08-588-985-2
15	172.5	2.4	1865	1	US-08-971-988-2
16	170.5	2.4	3878	4	US-09-914-259-11
17	170	2.4	1187	1	US-08-320-559-28
18	170	2.4	1187	3	US-08-545-860D-28
19	170	2.4	1187	5	PCT-US94-04496-28
20	170	2.4	1210	1	US-08-320-559-26
21	170	2.4	1210	3	US-08-545-860D-26
22	170	2.4	1210	5	PCT-US94-04496-26
23	168.5	2.3	3433	4	US-09-091-501B-10
24	168	2.3	564	3	US-09-308-022-6
25	166	2.3	947	4	US-09-418-780A-1
26	166	2.3	1231	4	US-08-714-741-41
27	165	2.3	1333	3	US-09-356-952-2

28	165	2.3	1850	4	US-09-620-093A-5	Sequence 5, Appl1
29	164.5	2.3	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
30	160	2.2	1312	4	US-09-345-882-29	Sequence 29, Appl
31	158.5	2.2	714	2	US-08-990-114-3	Sequence 3, Appl1
32	158.5	2.2	714	4	US-09-241-333-3	Sequence 3, Appl1
33	158.5	2.2	1507	4	US-09-914-259-37	Sequence 37, Appl
34	158	2.2	2474	4	US-08-265-967C-3	Sequence 3, Appl1
35	158	2.2	2474	4	US-08-305-790B-4	Sequence 4, Appl1
36	158	2.2	2972	4	US-09-579-181-2	Sequence 2, Appl1
37	158	2.2	3118	4	US-09-579-181-1	Sequence 1, Appl1
38	157.5	2.2	1111	4	US-09-914-259-28	Sequence 28, Appl
39	156.5	2.2	414	1	US-07-667-276A-4	Sequence 4, Appl1
40	156.5	2.2	1588	5	PCT-US93-07261-11	Sequence 11, Appl
41	156.5	2.2	1663	5	PCT-US93-07261-16	Sequence 16, Appl
42	154.5	2.1	631	3	US-08-847-065-25	Sequence 25, Appl
43	154.5	2.1	1848	3	US-08-296-791-6	Sequence 6, Appl1
44	154.5	2.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl1
45	154.5	2.1	2101	1	US-08-466-390-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-595-684B-31

Sequence 31, Application US/09595684B
Patent No. 6544766

GENERAL INFORMATION:

APPLICANT: Beraud, Christophe

APPLICANT: Ohashi, Cara

APPLICANT: Saksiewicz, Roman

APPLICANT: Vaisberg, Eugeni

APPLICANT: Wood, Kenneth

APPLICANT: Yu, Ming

TITLE OF INVENTION: Human kinesins and methods of producing

TITLE OF INVENTION: and purifying human kinesins

FILE REFERENCE: cytop036

CURRENT APPLICATION NUMBER: US/09/595,684B

CURRENT FILING DATE: 2002-06-24

PRIOR APPLICATION NUMBER: 09/295,612

PRIOR FILING DATE: 2000-04-20

NUMBER OF SEQ ID NOS: 105

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 31

LENGTH: 2662

TYPE: PRT

ORGANISM: Human

US-09-595-684B-31

Query Match

Best Local Similarity 18.2%; Pred. No. 2.3e-09;

Matches 274; Conservative 270; Mismatches 540; Indels 425; Gaps 64;

17 PGVKISIKISKEEVRRLKMYKPFMDODDSE-----EKELYLALHLASD 66

333 PYVNEVS---IDALLKRR---KEIMDKOLEVSLPRAQAMEKD---QLADLLEK 383

67 FFLKPGKDVRLVLAACLAIDFRIVAPAPYSPDKLIDF-MFTROLKGL-EDYKSPQ 124

384 DLQK-----VQNEKIEMLFRLVLTSSSLFLOGLKAKR 417

125 ENRYIYLLLENIAWVSYNICPELEDSNELFPOLYLTLSVINGNNOVHHMVDLMSSI 184

418 KRRVWPCGKIKMKNSN-----YADQFNIPN-ITTRTHKLSIMLREI 461

185 ---ICEGDTVSGELDPVL-VNLVPAHKLNKQAVDLAKALKLRTAQAIIEPYITFEFNOV 240

462 DESVSESDVFSNLTDTLSEIEMNPATLNDENIE----- 497

241 LMLGTSISDSEHVEDLILELYNIDSHLSVLPOLFEKLSNDNEERLQVVKLLAKMF 300

498 -----SELNSLRADVDNLVL-----DYQLRPREKEMELKKEKNDLDFEALERKR-- 545

```

QY 301 GAKSELASONKPLMOCYLGFRNDIHPVIRLECVKASHCMLMHPDLAKDLTEYLKVRSH 360
D 546 --KDOEQ-----LIHISNKNVKNVHREYVNO 571
QY 361 DPEAIRHVDIVSVTAAKKDILLVNDHLNFBVRERLDRKVRKREKMMGLADLYKKA 420
D 572 DLEBEL-----SSKVELLEKEKQIKLOEYIDSOX--LENIKMLOS-----YS 613
QY 421 LOSAGKDAKQIAMIKDKLHIIYONSIDRLVERIFAQYVWPHNLEETERRKCLYLL 480
D 614 LESIEDPKOMQULFDETV-----ALDARESAPFLSENLELKEKKELATY 661
QY 481 YATIDLAIVKALNEMKCONLRHQVLDLILQPKTDASVKAIFSKVMYITR----- 534
D 662 YKOME-----NDIOYQSOLEAKKMQYDLEKE-----LOSAPNEITKLSLIDKV 708
QY 535 -----NLDPGRKQODEMKKFTOYLEDDEKIRKOLEVLYSPTCGKQAGCVREITKLG 588
D 709 PKDILCNLEBEGKTIDQKELNKVEEVEALREEV-ILLS--ELKSLPSEVERLREI- 763
QY 589 NPKOPTNPLEMIKFLERLAPVHIDES-ISALIKOVNKSIDGTADEDEGVPTDOAI- 646
D 764 ---QDKSEBELHIIITSEKDKLEFSEVYHKESRYQGLLEBEGTKKDLATQSNYKSTDQEQ 820
QY 647 -----RAGLELIKVSFTHPLSFHSASFESTILACLMD----- 680
D 821 NFKTLHDFEOKYKMYLEENERNMOEIVNLSK-----EAQRFDSISGALKTELSTYKT 872
QY 681 ---DEKVAEALQJFKMTGSKIEEDPHIRSAALLPVLHKS-----K 719
D 873 QELOEKREVOERL--NEMEOLEKOLERDSPLQTVBEREKLITREKLOOLEEVKTLQOE 930
QY 720 KGPPOAKYALHCHIAIFSSKETQFAQIFELHKSIDPSNLEHITLPTVIGITALAPD 779
D 931 KDKDKLOESIQI-----ERDQKSDIHDVTVMNIDPQ--EQRLNLESKX-----Q 976
QY 780 QFAPKMSVATFVYKDLMDRLPGKTTKLVYDDEVSFETVAKIOALIMAVRWLIGM 839
D 977 ETTINTLSKISEEVSRLHM-----BENGTEDEFO-----QKMWGI 1014
QY 840 KNNHSGSTSLRLTLILHSDGLTEQGISK--PMSRLRLAAGSAIVKLAQ-- 891
D 1015 DKKODLEAKNT-QTLTADVKNETIIBOQRKIFSLIOEKNELOQMESYIAKEKLDK 1073
QY 892 EPCYHEITTEQYQCALAINDECYQYVYFAQRLHKSRLRLPLEYMAICALCA--KD 949
D 1074 ENIMETIENBELRL-----LQDELKQOEIYAQEKNAIKR--EGELSPCDRLAEVEE 1126
QY 950 PVKERRAHARQCLVKNINVRREYIKOHAAYSE-----KLSLLEEVYVPIYIHLAH 1001
D 1127 KLEKSSOQLOEKQOOLINVOEEMSEMOCKINEIENLKNELKNELTLEHMETERLELAQ 1186
QY 1002 -DPYVAVODI-EQLKDVKECLWFEVLILMAKNENSHAFIRKAVENIKQTKDAQDPDA 1059
D 1187 LANEYEEVKSTIKRKYLKE-----LOKSPETERDHLRGYIRELEATGLQIKELKLAHI 1241
QY 1060 KMNELTYVCDVANNITMSKSTYVLSLSPKDPVLPARFPTQPDKNFSNTKNVLPPEKSF 1119
D 1242 HLKEHOETIDELRRSV--SEKT-----AQIINQOLEKSHTK--LQEEIP-- 1282
QY 1120 FTGOKPKTTVNLGAVNKPSSAGKQOSQTS--SMEMEVSANASSNSSPKRLINGRDDS 1177
D 1283 -----VLHEOEELLPVYKVSFQETMNELELTLEOSTYTKDSTYIARI----- 1325
QY 1178 EMDHSE-NEDYTMSPLPGKSKDKRDSDLVRSLEKPRGRKKTTPVE-----Q 1225
D 1326 EMELERLNKEKFOESQOEIKSTIKRDMUKITIKALEVKHDKLKHITETLAKIOESOSKQ 1385
QY 1226 EERKLGW---DULTKLVQEQ--KPK-----GSGORSKRKH-----TASBSE 1261
D 1386 EOSLNMKEKNETKRIYSEMOQFKPKDSALLRIEIMJGLSKRLQESHDEMKSVAKEKDD 1445
QY 1262 QQ-----WPEERKLKEDI-----LENEDEONSPFKKGRKGRPP-KPLGGGTPKEEPT 1307

```

```

D 1446 LQRLQEVLOSDDQLEKKEIKVAKHLETEBEELKVASHCCLKEQDETINELVNLSKEKETE 1505
QY 1308 MKTSKK-----GSKKSGPPAPEEEDP---ERQSGNTEOKSKSGKQVRSAQORA-ES 1358
D 1506 ISTQKOLEAINDKLQNKIQEYIKERQOLNIQISEOVENVNELKQREHKKRADSALQS 1565
QY 1359 PESSAIEST 1367
D 1566 IESKMELEL 1574

RESULT 2
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 2.8%; Score 201.5; DB 1; Length 3248;
Best Local Similarity 19.5%; Pred. No. 1.2e-06;
Matches 298; Conservative 234; Mismatches 590; Indels 407; Gaps 63;

QY 19 VKESIDKI-----SKEMVRLKMWVTFMDMODSSEBEKELY----- 56
D 1900 MKELDSKLHIOEYQVLMIKIEACIELEKIVGELKKNESDLEKLEYFSCDHQELLQRYETS 1959
QY 57 --LNLALHLASDFLKHGKQDVLLVACCLADIFRIYAPAPYTPSPDKLDIFEMFITRQL 114
D 1960 EGLNSIDLEMHAD--KSSREDIGDNVA-----KVNDSWKERFLDYENEL 2000
QY 115 KGLDETKSPQFNRYFYLENIAMWKSYNICELEDSN--ELTFQLYTFLSPYINNGHNO 172
D 2001 SRINSEKASIEHEALYLEADLEVYOTKEKLEKDNENKOKYIVCLEBEL--SVYTSERNQL 2059

```

```

QY 173 VHMHWDLMSSTICBGPVSOELLDPVLNVLPAHKNLNKQAVDLAKALLKRTAALIEPY 232
Db 2060 --RGELDTWKSRTTLLDOLSEKMER--TQELSHQSECLHCIOVAEAVKERTELLQ-- 2113
QY 233 ITTFNOVLMKTSISDLSEHVPDL-----LELYNDSHLLSLVLPOLERKLSNDNE 287
Db 2114 -----TSSDVSSELLKDKTHQOEKLOSLKESQALSILTKCELENOIOLNKE 2160
QY 288 ERLQV--VKLLAKMFAGANDELSAONKPLMOCYL--GRENDIHVPRIECVFPASHCLMN 343
Db 2161 KELLKESSESLOARLESSEDPYELKLVNSKALEALVEKGEF-----ALRLSSTQEEVHQQLR 2215
QY 344 HPDLAKDLLEYLVKSHDPEEALIRHDVYSITYTAKKOLLVNDHLNVRERTIDKR-- 401
Db 2216 G-----TEKLRVIEADEKQOLH--IAEKLERERENDSLKDVENIERELOQSEENO 2266
QY 402 -----WVRKEAMMGLAQIYKKAQASAGKD--AAQJAWTKDLHIY 444
Db 2267 ELVILDAENSKAEVETLKTQIEMARSLKIFELDLVTLRSEKENLTKQIOEKQOGLSELD 2326
QY 445 YONSIDDLILVERIAQIYVPHNLFTTERMKCLYLYATLDLNAVAKALNEMKCONLRLH 504
Db 2327 KLLSFSKSLLEKEQEAEIOIKEESKTAVEM-----IQN 2359
QY 505 QVKDLIDLKOPKTDASVKAIFSKVMVITRNLDP-----GKAQDFMKFTQVLDDEKI 559
Db 2360 QKTELNEAVNALCGD-----QELMKATEOSLDPIREEHOLRSTIEKLRLAEDEK- 2411
QY 560 RKOLEVLVSPSCQKQAEQVREITKLGPNKOPNPFLMIKFLERITAPAHIDTESIS 619
Db 2412 -KOLCVLQ---QKLESEHMDLKGVRNENLRE---LEIRATQENHAA---LEENENK 2459
QY 620 ALIKOVNKSIDCTADEDEGVPTDQAIRAGLELLKVLSTFHPISHSAPESLLACLKM 679
Db 2460 GEVETLKAKIEGMT-----QSLR-GLEL-----DVTIRSEK--ENLNLNELOK 2499
QY 680 DDEKVAEAL--QIEKNTGSKIEEDFPHIR---SALLPVLHHSKKGPPROAKYALHCH 734
Db 2500 EBERISELEITINSSENILQOEKQVOKKESSTAMELOTLQLELNERA--ALHNDQ 2557
QY 735 AIFSSKETQFAOIFEPHL-----KSLDPSNLEHLITPLVYIGHIALAPQFAAPWKS 787
Db 2558 EACKKKEGNLSQVCLLEKRAQLOGLDEAKNNIYVLOSSKVG----- 2601
QY 788 WVAFTIYVDLMDNRLPGKKTITKLVNPDEEVSPEPMVTKIQAIKMVRMLGKNHNS--K 845
Db 2602 -----LIQEVEDGKQ--KLEKKDEEIS--RLKNQIODEQVLVSKLSQVEGEHQLMK 2648
QY 846 SGTSLRLTLTLHSDGLTEQKISKPDMSRLRLAAGSAIYKLAQEPCHHITLLEQY 905
Db 2649 EONELRLNLTVELOKIOVLOSKNASLD-----TLEVLQ 2683
QY 906 LCALAINDECYQV---ROVFAOKLHKGLSR---LRPLVEYMAICALCADDY---KERRA 956
Db 2684 SSYKLENEBELTTRKDKMSFYKVKAKTRAKETELQREHMEMAKTALDEELSGEKNRIA 2743
QY 957 HAROCLVKNINVRREYKLOHAHVSEKLSLAP-----EYVVPYTHILLAND 1002
Db 2744 GELQILLBEIKSSKQOLKELTLENSELKSLDCMHKQYKREELAEVQLRL--HE 2801
QY 1003 PD-----YKQVODI--EQLDVKECL---WFEVLIIMAKNE--NNSHAFIRK 1042
Db 2802 AEKKHQAALLDNTKOYEVEIQYREKLTSKKECLSSOKLELDLLKSSKEELNNSLKATQO 2861
QY 1043 MVENIKQTKDAQPPDAKMNKRLTYVCVA---ANNIIMSKSTYTESLESPKDPV-----L 1093
Db 2862 ILEELKTKR---MONLKYVNOGLKKNENRAGQKMKLLI--KSCKOLEEKEKTELQLELSL 2915
QY 1094 PARFTQPKNFSNTR--NLPPEMKSFTTPGKPKTTNV-----IGAVNKPPLSSAGK 1143
Db 2916 QAAQEKQKTGVTMDTKVDELITTEIKELKETLEKTKREADEVLDKYCSLLISHEKLEKAKE 2975

```

```

QY 1144 OSQTK-----SSRMEVSNASSSSNPSPGRIRKGRLDSDSEMDHSENEGYTMSSPLP----- 1194
Db 2976 METQVAHLCSQOQSKODSNGSLPLGVPVG-----PSPIPSYTE 3014
QY 1195 -----CKRSKDRDSDLVRSLELEKPRGRKTPYTEQEKLGMDLTKLVQEQKPKGSQ 1247
Db 3015 KRLSSGQNKASGKRQ-----RSSGIWENGRGPTPAT-----PESFSKSKS 3054
QY 1248 RSRKRG-RTASQSDQOPEEKRLKEDIENEDQNSPPKKGKRGPRPPPLGGGP--KE 1304
Db 3055 KAVMSGIHPA-----EDTEGTEFEPEGJLPEYVVKGFADIPTKTSPYILR 3099
QY 1305 EPTMKT---SKKSKKSGPPAPDEEERQSGNTQESKSKQHRVSRRAQRAESPSS 1361
Db 3100 RTTMAKTRPSRLAOKLALSPSLGKENIAESSKPTAGSRSQKVAQOR-----SPVD 3153
QY 1362 SAI-----ESTQSTPOKGRGPRKPTSPSQ 1386
Db 3154 SCTILREPTTKSVPVNLLPERSPTDSPRE 3182

```

```

RESULT 3
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1
;
; Query Match
; Best Local Similarity 2.8%; Score 201.5; DB 5; Length 3248;
; Matches 298; Conservative 234; Mismatches 590; Indels 407; Gaps 63;
;
; 19 VKESDKI-----SKENVRRLKMMVKTPMDMDQOSEEEKELY----- 56
; 1900 MKEIDSKLHDEVOQMTKIEACIELEKIVGELKKNESDSEKLETFSCDHQBLRQVETS 1959

```

QY	57	LNALHMLADDFLEHKGKQVRLLYACCLDIRITVAREPAPYSPDKIOTFETROL	114
Db	1960	EGLSNDELMHAD--KSSREDIGNVA-----KVNSMKSRFLDYENEL	2000
QY	115	KGLEDTKSPQNRREYLLLENITAMWKSVCICELEDSN--EIFTOLYRTFVINGHNOK	172
Db	2001	SRIBREKASIHHEALYLEADLEVJOTETKCLEKNOENOKIYVLEEEL-SVATISERNOL	2059
QY	173	VHMHVADSSITCEGDTVSOELLDTVULVNLVPAHKNLNKOAYDALAKALRTAOAIEPY	232
Db	2060	--RGELDMSKRTTALDOLSEKMEK--TOELESHOSCEHICIOVAEAEVEXKTELLQ--	2113
QY	233	ITTFENQVIMIGTISITPSLSEHVFELI-----LEYUNIDSHLLSVLPOLFETKXSDNE	287
Db	2114	-----TLSSVSELDKDTNLOEKLOSLEKDSOALSITCKELENQAOULKE	2160
QY	288	ERLOY--VKLLAKMEGANDSELASQNPWOCYL--GRNDIHPVIRLECVAFASHCLM	343
Db	2161	KELLYKESESIQARLESDEYEKLIWYSKALEALVEKGEF-----ALRSTOEBEHOJLR	2215
QY	344	HPDLAKDLTELKYARSHDPEAIIHDVVSUYTAKKIDILLVNDHILNFVARELIDLR--	401
Db	2216	G-----TEKLRVRIEDEKKOLH--TAEKLERERKNDSLKQVLESELMOSEBNO	2266
QY	402	-----WRRKEMAMGIAOIKKALASOAAKD--AAKOARIKIDLIHY	444
Db	2267	ELVILDAENSKAEVETLTQTOIEBMAKSLKIFELDLVLRSEKNMLTQIOKQOOLSBD	2326
QY	445	YONSTDRLLYERIFAQVWPHNLLETBRMKCLYUATLIDLNAVKALNEMKCONLIRH	504
Db	2327	KLSSFKSLLEEKDOALEIOIKESKTAVEM-----LON	2359
QY	505	QVKDLIDLINOKPKIDASVAKIFSKVMYITRMLPP-----GKADFMKKFOVEDEDEKI	559
Db	2360	QKLEINENVALCGD-----QELMKATBESIDRPRIEENHQLNISTEKLARLEADEK	2411
QY	560	RKOLEVULSPYSCSKOABGCYREITKILGNKPORTNPELEMIKFLLEIARIAPHIDTESIS	619
Db	2412	-KOLCVILO--OLKESHNADLKGRENLERE--LEIARINOEAAA--LEAENSK	2459
QY	620	ALIKOVNSIDGTADDEGEVPTOAIAGALEKLVLSFNIRISPHASEPESITLACIKM	679
Db	2460	GEVEIIRKIKITGM-----QSLR-GLEL-----DVVITRSEK--ENLTNELQK	2499
QY	680	DDEKVAEAL--QIFKNTGSKIEEDFPHIR--SALLPVLIHNSKKQBPPOAKAYALICIH	734
Db	2500	EOERISELEIINSFENILOKEOEKBOYMKREKSTAMEMLOTKELNEBVA--ALHNDQ	2557
QY	735	ALFSKKEQFOAIEPRH-----KSLDPSNLBHLITPLVTIGHIALNDQFALAPWS	787
Db	2558	EACAKKEONLSSOYCELEKLAOLLOGDEKKNNTIYLOSSVKG-----	2601
QY	788	WVATFIYVILMNRLPGKTTKTILMPVDEEVSPEWYIAQIKIMVEMILMKNNHS--K	845
Db	2602	-----LIGVEDEOKQ--KLEKDEDEIS-RLKNOIQOQEBDVLKSLGSVESEHOLMK	2648
QY	846	SGTSLRLITLTIHSDDLTBOGKISKPDMSRLYLAAGSALIVKLAQEBCHETITLEBOYQ	905
Db	2649	EONDELIRMLYTELOKIOVLONSKNASLOD-----TLEVIQ	2683
QY	906	LCALALINDECYQV--ROVFAOKLHKGLSR--LRLPREVAYICALAKDPV--KEBA	956
Db	2684	SSYKNLEBELELTLMKDKSPFEYKANKTKATLETLEORENHEMAOKTAELOEBLSEKKNLA	2743
QY	957	HAROCILKNINVRREYIKONAASEKLLSLP-----EYVVPYUITHLHND	1002
Db	2744	GELDILLLEIKSSDKOELTELSENSELKSLDCKHNDQVEKEGVAREIAUYOURL--HE	2801
QY	1003	PD-----YKVQDI--POLKDVKECL--WVLEIIMAKNE--NNSHATIRK	1042
Db	2802	AEKKNQALLDNTNKQOYEIOGTVEBKLTJSEKSSJOSKLEIDLKSSKEEELNLSYKATTO	2861

Best Local Similarity 19.4%; Pred. No. 2.8e-06;
Matches 297; Conservative 233; Mismatches 592; Indels 407; Gaps 63;

```

OY 19 VKELSDKI-----SKEENVRLKMYVTFMDQDSEEEKELY-----56
Db 1172 MKELSDKHLQEVOLMTKIEACIELEKIELEKKNESDLESEKLEFESCDHDELLORVETS 12331
OY 57 --LNLALHLASDFLHPGKDVRLVACCIAIDIFITVPEAVYTPDLKIDFMFTIROL 114
Db 1232 EGLNSDLEKHD--KSSREDGDVNA-----KYVDSWKREFFLDVENEL 1272
OY 115 KGLEDTKSPQFNRYFLLENIAWVSYNICELEDSN--EIFTQLYRTLFESVINGNHCK 172
Db 1273 SRIRSEKASIEHEALYLEADLEVOTFEKLECKENENOKVYVCELEEL-SVYTSERNQL 1331
OY 173 VHMHWADLMSSITCGDTVSOELDTVLVNLVPAHKNLNKAQYDLAKLLKRTQAQAIPEY 232
Db 1332 --KGLDTPMSKTTALDQISEKMEK--TOELESHQSECLHCIOVAEAEVEKEKTELLQ-- 1385
OY 233 ITTFNOYMLGKTSISDLSHEVFDLI-----LELYNIDSHELLSVLPQLEFKLKSNDNE 287
Db 1386 -----TSSDVSSELLKDKTHLOEKQSLSEKSOALSITKCELENOIAQLNKE 1432
OY 288 ERLQY--VKILAKMFGAKDSSELAONKPLMOCYL--GRFNDIHVPIRELCVYFASHCLMN 343
Db 1433 KELVYKESSEIQARLSESDYELKLVSKALEALVYKGF-----ALRISTOEVEHQILR 1487
OY 344 HPDLAKDILEYKVRSHDPEEAIHRDVIVSYVTAARKDILLVNDHLNLFVBERLTDK-- 400
Db 1488 G-----IEKRLVYIEADEKKQLH--IAEKLERERENDSLKDKVENLERELQSEBNO 1538
OY 401 -----RMVVRKEAMGIAQIYKKYALO-----SAAGKDAKQIAMIKKLHITY 444
Db 1539 ELVVIDAENSRAEVETLQIIEMARSLKVELDLVTLRSEKENLTQIOEKQGLSELD 1598
OY 445 YONSIDRLVERIFQAYVPHNLLETTERMKCLYLYATLNLNAKALNEMKCONLRLH 504
Db 1599 KLSSEFKSLBEKEBAELQIKESEKTVEM-----LQN 1631
OY 505 QVKDLDLILKOPKTDASVKAIFSKVAVITRNL.PDB-----GRAQFMKFTVOLEDEKI 559
Db 1632 QKELNEVVAALCGP-----QETMKATEQSDLPRIEENHQLRSIKELRLARLADER- 1683
OY 560 RKOLEVLVSPYCSCQOAGCVREITKKGKGNKQPTNPLFMKLELEIAAPHIDTESIS 619
Db 1684 -KOLCVLO-----QKSESHADLKGRENLERE--LETARTQOEIAA--LEAENSK 1731
OY 620 ALIKOVNKSIDGTADDEDEGVPTQAIKAGLELKLVSFTPHISFHSASEFESLACIKM 679
Db 1732 GEVELLAKKIEGMT-----QSUR-GIEL-----DVTITRSEK--ENLTNELOK 1771
OY 680 DDEKVAEAL--QIFKNTGSKIEEDFPIH--SALLFVLNHHKSKGPPROAKYAHCIH 734
Db 1772 EOERISELEIINSSEENILOKEOEKVMKEKSSPAMEMLOQKELNERVA--ALHNDQ 1829
OY 735 AIFSKKEQFAQITEPLH-----KSLDSPNLEHLITPLYTIGIALALADQFAAPKKS 787
Db 1830 EACRAKKEONLSSOVCETELEKAQLOGLGDEAKNNIYVQSSVNG-----1873
OY 788 WVAFTVYKDLMLNRLPGKTTKTLVMPDEEVSPTMYVIAQIKMVAVRLKMNKNNHS--K 845
Db 1874 -----LIGVEEDGQ--KLEKKEDEIS-RIKNDIOQOEQLVSKLSOVEGHQJWK 1920
OY 846 SGTSTLRLLTTLHSDDGLTEOGKISKPFMSRLRLAAGSAIVKLAQEPCHYEITLLEOYQ 905
Db 1921 EONLELRNLVLELEQKIOVLQSKNALSOD-----TLEVLQ 1955
OY 906 LCAIALINDECYQV--ROVFAOKLHKGLSR--LRPLEVMAICALCAKADPV--KEERA 956
Db 1956 SSYKNIENLELTLMDKMSFVEKYNAKMTAKETELQREHMAOQTAELQOEELSGEKNFLA 2015
OY 957 HAROCLVKNINVRREYLKQHAAVSEKLSLTP-----EYVVYTTIHLALND 1002

```

```

Db 2016 GELQLLLEBEIKSSKQDLKELTLENSELKSLDCKMHKQDVEKREKRYEBEIAEYQLR--HE 2073
OY 1003 PD-----YKQVDI--EOLKDYVECL--WFVLEIIMAKNE--NNSHAFIRK 1042
Db 2074 AEKHOALLDPTNKQYEVEIQTIRKILTSKECLSSORLEIDILKSSKEELNNSIKATQO 2133
OY 1043 MVENIKOTKDAQGGPDADAMNEKLYTVCVA--NNIIMSKSTVYLSLEPKDPV-----L 1093
Db 2134 ILEBELKTK--MDNLKYVQOLKKEKNERAAGKAKLLI--KSCQOLEEKEITLQKELSOL 2187
OY 1094 PAREFTOPDKNFSNTR--NYLPPEMKSPFTPKPPTTVN-----LGAVNKPLSSAGK 1143
Db 2188 QAAEQKQKGTVMQTKVDELITTEIKLEKTELEKTKADEYDLKYCSLLISHKELEKKE 2247
OY 1144 QSOIK-----SSRRETVSNASSSNPSSPGIKRGLDSSEMDHSENEDEYITMSPLP-- 1194
Db 2248 MLETOVAMLCQOQKQSDSGSPLGLPVVPG-----PSPDIPSVTE 2286
OY 1195 -----GKSKDKRDDSDLVRSLELEKPRGKKTPTVBOEKLGMDDLTKLVQEQKPKGSO 1247
Db 2287 KRLSSGONKASGKHO-----RSSGIWENGSGPTPAT-----PESFSKSK 2326
OY 1248 RSRKRG--HTASESEDOQWPEEKRLKEDILENEDQNSPPKKGRGPRPKPLGGCTP--KE 1304
Db 2327 KAVMSGIHPA-----EDTEGTEFEPEGLEPVYVKKGFADIPGKTSPTILR 2371
OY 1305 EPTMKT--SKKSKKSGPPAPEBEEEROSGTEOKSKSKOHRVSRRAQRAESPES 1361
Db 2372 RTVMATRTSPRLAOKLALPSLISLGEKMLAESKPTAGSGRSOKVYKVAOR-----SPVD 2425
OY 1362 SAI-----ESTOSTQKGRGPRSKTPSPSQ 1386
Db 2426 SGTILREPTTKSVPVNLPERSPTDSPRE 2454

```

```

RESULT 5
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

Query Match 2.6%; Score 190; DB 4; Length 10182;
Best Local Similarity 19.1%; Pred. No. 6.8e-05;
Matches 276; Conservative 235; Mismatches 523; Indels 414; Gaps 67;

```

OY 23 SDKISKEEMVRLKMYVTFMDQDSEEEKELYLNALHLASDFLHPGKDVRLVAC 82
Db 8944 NDEITMNTNSIAOLSRIVQAFAFLDLDANKSLDELNNOAFVQAASSNYINSDCLKQOFDI 9003
OY 83 CLADIFRIYAPEAAYTSPDKLIDFMFTIROLKGL--EDTKSPQFNRYFLLENIAWV 138
Db 9004 ALSNARKVLEKNEKNKNDK-----QIOGLKQVIEDTKO-----9037
OY 139 KSYVICELEDSNEITFQLYRTLFESVING--HNQYHMHMVDLMSSI-----ICEGDT 190
Db 9038 -ALNGIORLSKAKAKAIQYVOSTL-SYINDAQRIHAENNIHNSDLSLAWTLKASADLDN 9095

```

```

QY 191 VSOELDTV-----LVNLVPAHKNLN---KQAVDLAKALLKRT----- 225
Db 9096 AMDLRBTIESNSTVSPNSVNYINADKNQIEFDEALQASASIKSENPATIEEVLGL 9155
QY 226 AQAI-----EPITTFNOVLMGKTSISDLSEHVDLLELYNIDSHLLSVLPQ 276
Db 9156 SQAIYDTKNALNCEORLATEKSKDLKLIK-GLKDLNKAQLEEDVTN--KVSANTLLELSQ 9212
QY 277 L-EFKLSNDNEERLO-VVKLAKMKGAK-----SELASO--NKPLMOC--YLGRENDI 325
Db 9213 LTOSTELNDKMKLTKLTKTLVNPVKASLNYRNADNLKROFNKALKEKAGVLLNKSGT 9272
QY 326 HVEIRLECVAFASHCLMNHDPKAD-LTEYLKVRSHPEBAI-----RHDIYSIVTA--- 377
Db 9273 NVNI-----NDIOH-LITQIDNAKQDQNGERRLKEHQKSEVFLIKELDILNNAQKAI 9326
QY 378 -----AKDILLVDHLNFEVRETLDKRMVREKEMGLQITKTAALQ--SAAGKDAKQ 432
Db 9327 NOIRASDKITINOYDNAME-----LNDAMQGL-----KEHYAQLTATTKD---- 9368
QY 433 IAWIKDKLHIYONSIDDLLEVERLPAQYVPHNLE-----TEEMKCLYLYAT 483
Db 9369 -----NIEYLNADDEHKL-----QYDYAINLANNVLDKENGINKANITIGTION 9413
QY 484 LD-----LNAVKALNE--MMKQNLRRQVLDLILKQPTDASVKAIFSKVAVITENLP 537
Db 9414 MDAARALLNGIERLKAQOTAKHANDIKDTLRQDEIHNATWSKSA-QAKQWY----- 9466
QY 538 DPGKADFMKFTQVLEDEDEKIRKOLEVLSPTCS--CKQA--EG-----CYRETTK 586
Db 9467 -----NEARVALSNINDATSNLDVNAQKQEGOSAIEHINADELPA 9508
QY 587 LGNPQPTNPFELEKIKELERIAVHIIDTESISALIKQVKSIDGTADDEGVPTQAI 646
Db 9509 KIDANQMDQKVEDINHLISQ--NPNLSENEKKLISQINKLVINGINELQAIKQOIE 9566
QY 647 RAGLELLKLVSTFHPISTFSAETFEESLACLKMDDEVABAQOIFKNTGSKIEEDPHI 706
Db 9567 NATTKDEVIETETTKLIIAKAE-----AKOMIKELSOKKRD----- 9602
QY 707 RSLALPLVHLHKSCKPPOAKVAIHCHIAIFSSKETQFAQIFELHSLDP--SNLEHIL 764
Db 9603 -----AINNNTDLTPSOKA-----HALADIDKTEKDALOHEN--SNSIDIDINNKCHAF 9650
QY 765 TPLVITIGHIALAPDQ-----FAAP-----MKSVAATFIV-----KDLAMDRL 803
Db 9651 N---TLAHITITWPTDQPLVEFVPELSQNALVTSSEVVARDETISLESIKKKTILDEL 9707
QY 804 PGKKTTLKLVDP--EVSSEPTMKIQAI-----KMYRWILGKNHNSKSGTS 849
Db 9708 ---KVNIVSLPNTDKVDHLTAKVYVILADGSYTVVNPVKVVEKEIQIAKKDAIKTIDV 9764
QY 850 TLRLTLTIHSDDLLE-QGKISKRPMSRLRLAAGSAIVYLAQPCHEIIT----- 900
Db 9765 LVROKIKIDISNNELISTOREDAKAELEKIK--KQAIIDVNNHKSIKIDLEYVAKRTDFEE 9821
QY 901 LEOYOLCALAIN-----DECYQVROVFAQ--KLHGLSRLRLPLEYMAICALCAPDV 951
Db 9822 IDGFDPKRFITLNAKAKKIITDVMTQIONGKEIETIGLISNETQTDKQOLTAQKFELE 9881
QY 952 KERRAARQCLVKNINVRREY-----LKQHAASVEKLLSLPEYVVPYTHILLANDPD 1004
Db 9882 KVEHAHN---LVEINQOQEEFNRRYKHILNQAHLLGK-----HIAEKKLG 9924
QY 1005 YVAVQODIEQKDVKECLMFILE-----ILMAKNENSHAFIRKAVENIK 1048
Db 9925 YVAVNKTQOILINNOASAFYIKQWALDRIKOILETMNSIRGAHTVQOVHAKALLQIGIOIL 9984
QY 1049 QTKDA-----QCPDAMNEKILTYCDVAMNITISKSTYTESPDPV 1092
Db 9985 KVNYSIINOSFNDLSHNFYLSKFDARLEK-----DVANHIVQETETFEVYLGK--TGV 10037
QY 1093 LPARF--FTQPDKNFSNTKNVLPPEKMSFF--TPGKPKRTTNVLGAVNKPLSSACKQSGTK 1148

```

```

Db 10038 EPCKINKETQOQPLKHNNDLSLFKHLVDNFNGKTVGVIITLTGLISFVLAKRRKEE 10097
QY 1149 SSRMEFVSNASSSSNPSGRIKGRIDSSSEMDHSEMEDYTMSPLPBGKSSDKRDSLDYR 1208
Db 10098 KOSIK-----NHHKDIRLSD-----TDKIDPIVITK 10123
QY 1209 SELEKPRGRKPTVTEOEKLGMDLTKL--VOEQKPKGSORSRKRGHT-----ASESD 1260
Db 10124 RKIDK-----EQIQNDKHSIPVAKKHSKSKQKSEEDIHSPVKKRQNSD 10171
QY 1261 EQQPEEK 1268
Db 10172 NORXTEE 10179

RESULT 6
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Biles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swatlo, Edwin
; APPLICANT: Yoheer, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-0712
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8991 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; US-08-714-741-32

Query Match 2.68; Score 185; DB 4; Length 8991;
Best Local Similarity 20.08; Pred. No. 0.00014;
Matches 197; Conservative 120; Mismatches 322; Indels 348; Gaps 46;
QY 538 DP-GKADFMKFTQVLEDEDEKIRKOLEVLSPTCSQKQAGCVREITTKLGMPKQ-PTN 595
Db 8210 DPGKTDDELDKREA-----BAELNKVLEALPNQVAALEIE-----ELSKLEDLNKAQETN 8259

```


QY	596	PFLMFKRLERIPAVHIDTESIS	---ALIQVKSIDGT---	ADDEDGCVPIDQAIR	647
Db	8260	HVEDYIEKGLE	-----BAIQKQELKERTPELDALNLTGDPDDEETPAPEA-	8308	
QY	648	AGLEILKLSTPHIFISFAETFESILACLKMDKEVAE	-----ALAQIFKMTGSIKIEDF	703	
Db	8309	-----	PAE-----QPREKPAEETPAAPAPREKSADQQAEDY	8341	
QY	704	PHIRSAILLPVLIHKSCKGPPROAVYAIHCIIAIFSSKETQFAOJIEPLHKSIDPSNLEHL		763	
Db	8342	AR-RSE--EEYNRLTQOOPKAEKPA	-----PAPAPKEQA-----PAPKKQKQVNIENL	8389	
QY	764	ITPLVTYIGHI	-----AL-----LAP--DQFAAPKSWATFI	793	
Db	8390	LSTLDPGGTODELDKGAEAELNKKVEALPNPYXELSEELSPEDNLUKAETNHVEDYI		8449	
QY	794	VK-----	---DLMDRLP--GKTKTKLVN--PDEEV-----	818	
Db	8450	KEGLEEALATKQALEETPOEVDALNDLVYDGESEETPAPAPQDEPAPAPAPAEOPA		8509	
QY	819	---SPETMVKIQAIKMYRWLLGKNNHKSSTLRLTYLILHSDG--LTEO--GKI	870		
Db	8510	PAPPEKSADQQAEDYAR	-----RSEGYNRLTQOOPKA	8545	
QY	871	SKP-----	---DMSRLRLAGSAIYKLAOEPCYHETIT--LEQYOLCALAI	911	
Db	8546	EKPAPAPAPKEOPAPAPNKIEARIQ	-----SDLKAEENNVEDYIEKGLEQO-----AI	8594	
QY	912	NDECTQYV--ROVFQOKLHKGSLRLPLLEVYALCALAKDPV--KERAHACQLVKNINV		968	
Db	8595	TNKAELATQOINDKQKDELEDELEKY	-----LATLDPEGTODELDKGAEAELNE	8650	
QY	969	RREYLKQHAANVSEKLLSLPEYVVPYTIHLANDPYYKQODIEQLDKVCKCLFVLEITL		1028	
Db	8651	KVEALQOVALELEELS	-----KLED--NLKDAE-----	8677	
QY	1029	MAKKENNSHAFIRKAMVENIKOTKQAQCPDAAKMKIEKLYTCVDVAMNIIIMSKSTYSLESP		1088	
Db	8678	---TNNVEDYIEKGLEEALATKKAE	-----LEKQKELDAALINELGPDGDEETPA	8726	
QY	1089	KDPLVPARFPTQPKNFSNFKNYLPPEKKSFTFGKRTNVLAIVKPKLPSACKOSOTK		1148	
Db	8727	A-----POPEKPAEPEENPAP	-----APPEKSADQQAED	8757	
QY	1149	SSRME-----	---TVSMSSSSSNPSSPGRIKGRLDSEMDHSENDYT-----	1188	
Db	8758	YARSEEEYNRLTQOOPKAEKPAAPAPQEPAPAP--KIELKEIDESSEDYANEGFRAP		8816	
QY	1189	MSSPLPCKK-----	---SDKRDSDLVRSLELPRGKRTKPTVTEQEBEKLGMDLTKLYQ	1239	
Db	8817	LHSLRLDKKKKLSLEELSDKIDELD	-----AEIKLELDQLKAVENNINVE	8862	
QY	1240	EOKKGSORSKRRGHTASESDQOOWPEEKLKEDILENEDQNSPPPKGKGRPRKPLGG		1299	
Db	8863	DYIEGLEKLTIAAKTLEKTEAD	-----LKKVNEPEKSAEP-----SQEKPREE	8910	
QY	1300	GTPKEEPTMTKSKGSKSGAPPAPEEEEBEEROSGNTQOKSKSKOHVSRRAQORASP		1359	
Db	8911	APAEQGTLETPQEPKAEETPAPAPKEKPAEDQPNAKETDDQQAEDYARRSEEEYNRL	---	8967	
QY	1360	ESSAIESTQSTPOKGRGRPKSTPSPQ		1386	
Db	8968	-----TQOOPKAE--KPAAPQEPQ		8986	

RESULT 7
US-08-769-309A-5
Sequence 5, Application US/08769309A
Patent No. 571890
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauk, Theresa M.

1 TITLE OF INVENTION: Protein Binding Domains of Gravin
2
3 NUMBER OF SEQUENCES: 24
4
5 CORRESPONDENCE ADDRESSES:
6 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
7 STREET: 6300 Sears Tower/233 South Wacker Drive
8 CITY: Chicago
9 STATE: Illinois
10
11 COUNTRY: United States of America
12
13 ZIP: 60606-6402
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.30
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/769,309A
24
25 FILING DATE:
26
27 CLASSIFICATION: 435
28
29 ATTORNEY/AGENT INFORMATION:
30 NAME: No. 5741890and, Greta E.
31
32 REGISTRATION NUMBER: 35,302
33
34 REFERENCE/DOCKET NUMBER: 27866/33451
35
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 312-474-6300
38
39 TELEFAX: 312-474-0448
40
41 TELEX: 25-3856
42
43 INFORMATION FOR SEQ ID NO: 5:
44
45 SEQUENCE CHARACTERISTICS:
46
47 LENGTH: 1780 amino acids
48
49 TYPE: amino acid
50
51 TOPOLOGY: linear
52
53 MOLECULE TYPE: protein
54
55 JS-08-769-309A-5

Query Match	2.5%	Score 181	DB 1	Length 1780
Best Local Similarity	20.9%	Pred. No.2.le-05		
Matches	99	Conservative	59	Mismatches 151; Indels 164; Gaps 19
QY	1036	SHAETIRKVENIKOTKDAQDPDADKMKNEKLTVCDDVANNIIMSKTYSLSLSPKDPVLA	1095	
Db	247	SHAETSPRAESQGVAECKKEEGEEK	288	
QY	1096	RFTFQDPDKNPSTKNYLPRPEKSPFTPC	1144	
Db	289	-----SETES-----TFKKKFTQGMAGRKKTSPRKKKEDEVEASEKKKEDEPERKV	334	
QY	1145	SQTKSRMETVY-----NASSSSNPSSPG-----RIKGRLDSEMDHSENYDTMS-----	1190	
Db	335	DTEEDGDAEVAASEKLTASSEQAHQDPREPAESAHNEPRLSAVEKVELPSEEQVSGSQSPSEK	394	
QY	1191	SPLPGRKSKDRD-----SDLVRSLEKPRGRKKTPTYE-----QOEKLGMD-----	1233	
Db	395	PAPLATEVEDEKIEVHOEEVAEVAEHNVSVEERTEOKEVEETAGSVAEELVGDMAEPQ	454	
QY	1233	-----DLTKLVQ-----EOKPKSORRKRKGHTASESDQOWPEKRLKE-----	1272	
Db	455	EAEPAKELVTKKEFCVSGEDPTQGADLS-----PDEAKLSRPECVSEV	499	
QY	1273	DILENED-----QNSPPK-----KKGGRPKPLAGGT-----KKEEPTMKT	1311	
Db	500	EMLSOSQRMKQVQSPKLKLTFTSTGLKLSGKKOKKRGCGDESEGHYQVPADSPDSOE	559	
QY	1312	KKGSKKKSGPAPAE-----EPEEROGSONTFOKSKSKO-----	1343	
Db	560	QKGSSASSPPEPEEITCLEKGLAEVQODGAEAGCATSDGEKKRGGVTPMAAFKMMVTPK	619	
QY	1346	HVWSRRAO-----QRAESPESALISTOSTPOKGRGRPSKTPSPPOPKNV	1391	
Db	620	KVRAEPESDKEDLDLVKSAITLSTSESTASSEMDEMKGVSEEPKPEEPKRV	672	

RESULT 8
US-08-994-570-5

: Sequence 5, Application US/08994570
: Patent No. 6090929
: GENERAL INFORMATION:
: APPLICANT: Scott, John D.,
: APPLICANT: Nautel, Brian J.,
: APPLICANT: Klauke, Theresa M.,
: TITLE OF INVENTION: Protein Binding Domains of Gravin
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borum
: STREET: 6300 Sears Tower/233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/994,570
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 6090929and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/33451
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: TELEFAX: 25-3856
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1780 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-994-570-5

Query Match 2.5%; Score 181; DB 3; Length 1780;
Best Local Similarity 20.9%; Pred. No. 2.1e-05;
Matches 99; Conservative 59; Mismatches 151; Indels 164; Gaps 19;

QY 1036 SHAFIRKVVENIKQTFKDAQGDPAKMNKILTVCDVAMNIMSTYSLSPDPVLA 1095
DB 247 SHAETSPAESGQVAEECKEGERK-----QEKPSASPSPTSPV-- 288
QY 1096 RFTQPDKNFSNTKNYLPPEKKSFFTPG-----KPKTNVLGAANKPLSAGKQ 1144
DB 289 -----SETGS---TFKKFTQGMAGMRKTSFRKKEDEVEASEKKKQDEPKV 334
QY 1145 SQTSSRMETYS---MASSSSNPSPG-----RIKGRDSSSEMDHSENEITYMS----- 1190
DB 335 DTEEDGAEVAASEKTLTASEQAHPESAHEPRLSAEYKVELPSEBOVSGSOGPSEK 394
QY 1191 -SPLPKKSDKRD-----SDLVRSLEKPRGRKKTPTVE-----QEEKLMD----- 1232
DB 395 PAPLATEYFDEKIEVHOEYVAEYHVSTVEERTEDQKTEVETLGSVPAAELVGDAEPQ 454
QY 1233 -----DLTKLVQ-----EOKPKGSORSRRKRGHTASESDEQWPEEKRLKE----- 1272
DB 455 EAEPKELVKLKEFCVSGEDPTQCADLS-----PDEKLSKPPGAVSEV 499
QY 1273 DILENEDE---QNSPPK-----KGRGRPPKPLAGGT--PKEPTMKTS 1311
DB 500 EMLSSQERMKVQSGPLKLTSTGLKLSGKRGKRGSGDESGEHTQVADSPDSQEE 559
QY 1312 KKGSKKSGPPAPE-----EEEEROSGNTQKSKSKO----- 1345
DB 560 QKGSSASSPPEPFIITCLEGLAEVQODGAEGGATSDGKKKREGVTPMAFSKMWTPK 619
QY 1346 HRVSRRAQ-----QRAESPSSAISTOSTPOKGRGRPSKTPSPSPQKKNV 1391

DB 620 KRAVRPSESDKEDELDKVKRSATLSTESTASEQJEGKSGVEEPKPREPKRV 672

RESULT 9
US-08-198-446B-15
: Sequence 15, Application US/08198446B
: Patent No. 5674996
: GENERAL INFORMATION:
: APPLICANT: Hartwell, Leland H.
: APPLICANT: Weinert, Ted A.
: APPLICANT: Pilon, Sharon E.
: APPLICANT: Groudine, Mark T.
: TITLE OF INVENTION: Cell Cycle Checkpoint Genes
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
: STREET: 1420 Fifth Ave., Suite 2800
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101-2347
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/198,446B
: FILING DATE: 18-FEB-1994
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Shelness, Diana K.
: REGISTRATION NUMBER: 35,356
: REFERENCE/DOCKET NUMBER: FHCRI7537
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-682-8100
: TELEFAX: 206-224-0779
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2368 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: DESCRIPTION: Yeast MEC1 protein
: US-08-198-446B-15

Query Match 2.5%; Score 178.5; DB 1; Length 2368;
Best Local Similarity 19.4%; Pred. No. 5.5e-05;
Matches 251; Conservative 188; Mismatches 477; Indels 377; Gaps 60;

QY 68 FLKHPGKD--VRLIVACCLADIFRIYAPPEAPYTSPPDKLQDFMTITQLGLGETKSPQF 125
DB 526 FLSPHNIDEPSESLISLIFSLHRTFS---HQPPRLTQGNQINSFY-LVQKCFMNS 560
QY 126 NRYFYLLLENIAWVSYNI--CFELEDSENEIFTOLRPLFSYINNGHQAQKMHMVDLMSS 183
DB 581 NRYLRLL-----STRIILPLFNSDSHNSDEHTALILFLOSQKLPVKNENIYAWTQ 633
QY 184 IICBGDTVSQELDTVLNVLVPAHK-----LNKQAYDLAKALKLRQAQALEPYITP 235
DB 634 LTL---TTSNDVFTLLKLILDFNSDDYSLRIWMTLQIKMKATILKKTPEYQLLSPIPV 690
QY 236 FFNOYMLGKTSIDSEHVFDLIELXNDSHLLS-----VLPQLEFKLSMDNERL 290
DB 691 LLRQ---LGNLVE--RKVGQNLIELLGYPSKTIIDIFQRYIIPVAIIDYKSD----- 739
QY 291 QVVKLAKMFGAKSELSAONKP-----LWOCYDGR---F 322
DB 740 -VLSEIRKINQDGTSLINQMKVLLKKNRQIFAVALVYKHGLEFSDIILETLPLNRAPTF 798
QY 323 NDIHVPIRLBCVK-FASHCLMNHDLAKD-----LTXYLKVRSR--- 360

```

Db      739 DKGITAVLPDYKTLAEITKLYKNSVTKDASDENANMILCSLRFILTNFEKDKRHGSKY 858
Qy      361 -----DPEAIR---HDVIVSVTAAKRIDILYNHDLNFVEERTIDK-----RRRV 404
Db      859 KNINMTDQEQAFKKLQDNILGIFQVSSDIHVEGRTTYEKLRYINGISFLIYAP 918
Qy      405 RKEAMMGIAQIYKKAQSAAGDAKDAKJAMIKDKLHLYYONSIDRL--LVERIFA-- 460
Db      919 KKSIIISALAQI--SICLOTGLG---LKEVRYEAFKCMHLVHNLHDEELSTYIDSLIAFI 973
Qy      461 -----QYVVPNHEETTERMKCLYYL-----YATLDLNAVALNEMMKQNLRL-- 503
Db      974 LQKMFSENGKLNINYYSTIDPLIKREKSDILKIKPTTYLALYKRELIGTLANDGGFARVY 1033
Qy      504 HOVKRLDLIKOPKIDASKAIFSKVWYITRNLP-----PGKAQ 543
Db      1034 NKIRSTDPLI--PIFANNIKS--SNKYYINOLADIEVYLRRKQTERSIDFTPKKVGQTS 1089
Qy      544 DFMKFQTOVEDEDEKIRKQLEVLVSPQSCQAECCVREITPK-----LGNPKQPTNP 596
Db      1090 DITLVGLALDLSHKFNRDLKDLCEKCAKISMIG-VLDYTKHEKRRTTYSENEVYDLND 1148
Qy      597 FLEMIFELL---ERIAVPHIDTESIS-----ALIKQVNSIDGTADD---EDEGVPT 642
Db      1149 SVQTIKFLIWIINDILVPAFMOSENPQKQLEVALYIOESLKYCGLSSEMDNHKELYPN 1208
Qy      643 DQATFAGLELLKVLST--HPI--SFHSEFTESILACLKMDERYAER-----ALQ 690
Db      1209 EAKL--WEKNSVSKTITYPLTSSLYLAQSWKEVP--LKYPNNFKGQYIWKRRFTLD 1264
Qy      691 IFKNTG-----SKIEEDPHIRSALEPLVH-----HKSKGPP----- 723
Db      1265 LTK-TGTENPHGHVFSLLIREDDGSLNFIPLYSIDIIIAKEGCTPYADLNGIIEF 1323
Qy      724 -----ROAKYAIHCIAIFSSKE-----TOFAQIEPLH-----KSLDSPN 759
Db      1324 DSIFTCNLEGMNNLOVDSLRCMYESIFVFEYECCKWATEFKQYNSKLHGTFTIKDKTKTN 1383
Qy      760 LEHLTPLVITIGHITALAPDOFAAPKSWVATFYIKDLMDNRLP----- 804
Db      1384 M-----LHIDEFLRTTPS-----DLAORSLTDSFERSALYLEQCY 1421
Qy      805 -----GKTTKLWVPDEEV-----SPETWYIOAIKKMVRW----- 835
Db      1422 RQNPDRKQNGOLKLNQIITYEEIGDISLQGVLTFTKGNLVSITEELQISENKKLAQD 1481
Qy      836 ---LIGKNNHKSSTGLR-----LTTILHSDGDLTEQGIS--KPDMSRLAAGSA 885
Db      1482 CFNVIGKFSDDPKTTTRLKSMYDQOLYSQIISNSSFSHSDPKISLSPDVK----- 1532
Qy      886 IVKLAQEPCHIEITLBOYQCALAIND--CYQVQVPAQKIH--KGLSRRLPLEY--- 939
Db      1533 -----EWSISGLEANLEGNVOTLKNVQESLRLNIDREVLQYVIA 1576
Qy      940 MAICALCAKDPVKERRAHAROCLVNINVRREYLKQHAASSEKLSLIPYVVPYTIHL 999
Db      1577 KALAIASMDPLRQ-----KYINSRFLIGTNITTSKETTLLKQNLMLKMLSL 1627
Qy      1000 AHPDPYV-KVODIEOLKDVKECLWFLEILMAKKNNENSHAF--IRRMVENIKQTKDAQSPD 1057
Db      1628 -YDLSFLSSAKDKKEKSNNTYILDYRMERIGADEVPNHYILSMRKSQDLKMNDEA---- 1682
Qy      1058 DAKMNEKLYTVCDVAMN-----IIMSKSTYSLE 1086
Db      1683 DADIGKTFITLAQARNNARLDIASESLMHCL 1715

```

```

RESULT 10
US-08-870-693-15
; Sequence 15 Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.

```

```

; APPLICANT: Weinert, Ted A.
; APPLICANT: Pion, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCRI10798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2368 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: yeast MDC1 protein
;
; US-08-870-693-15
;
; Query Match 2.5%; Score 178.5; DB 2; Length 2368;
; Best Local Similarity 19.4%; Pred. No. 5.5e-05;
; Matches 251; Conservative 188; Mismatches 477; Indels 377; Gaps 60;
;
Qy      68 FLKHPGKD--VRLVACCLADIFRIYAEAPYTSFDKLDIFMFTTRQLKGLDPTKSFQF 125
Db      526 FLSHNIDFSESLSGLFSLHRFS---HFQPKLTDNGQINKSFK-LVQKCFMNS 580
Qy      126 NRYEFLLENIAVMVSYNI--CFELEDSEIFTOYRTLFSVINGNHQAHHMHWDLMS 183
Db      561 NRYELRL-----STRILPFIINSDSHNSEDEHTATILIKLQSKKLEPVVENLYAMTQ 633
Qy      184 IICEGDIVSOELDLTVLVNLPVPAKN-----LNKQAYDLAKALKRTQAIEPYITTT 235
Db      634 LTL--LTSNDVFDPLTLKLDIFNSDDQSLRIMWTIQIKMKAKILKTPQLLSPILPV 690
Qy      236 FFNQVLMGKISISDSLSHVVDLILELYNIDSHLLS-----VLPOLEFKLSNDNERL 290
Db      691 LLRQ---LGNLVE--RRVGFQNLIELLGYPSKTYLIDIFORYIIPYALIQKSD----- 739
Qy      291 QVVKLAKMFGAKDELSAQNKP-----LMQCYLGR---F 322
Db      740 -VLSEIAIKMGDQPSLNLQMKVNLKKNRQIFAVALYKGLSLDLELFLNRPATF 798
Qy      323 NDHVPRLKLVK--FASHCLMNHPLDAD-----LLEYLKVSH--- 360
Db      799 DKGITAVLPDYKTLAEITKLYKNSVTKDASDENANMILCSLRFILTNFEKDKRHGSKY 858

```

```

QY 361 -----DPEAIR---HDVIVSYTAAKKDILLVNDHLLNFVRETRTDK-----RWRV 404
Db 859 KNINMTDDDOEQAOFQKLDQNLIGIFOVESSDIHVGCRTTYEKLKLVINGISPLIYAP 918
QY 405 KREAMMGLAOYKRYKALQSAAGDAKDAKOIAMIKDKLHIYQNSIDRL--LVERIPA-- 460
Db 919 KKSIIISALAOI--SICLOTGLG--LKEVREAFRCNHLVRLHNLDELSTVIDSLAFI 973
QY 461 -----QYVVPINLETTEKMKCLYLL-----YATLDLNAVKALNEMKCONLRL-- 503
Db 974 LQKWFSENGKLRLNIVYSIDTLTKESDILLKLPYTTALVGPBELGILARDQFARMV 1033
QY 504 HOVADLDLJKOPKTDASVAFKSVAVITRNLPD-----PGKAQ 543
Db 1034 NKIRSTDIL--PIFANNLKS--SNKYVINQNLDDIEVYLRKOTERSIDETPKKVGOTS 1089
QY 544 DFMKFTQVLEDEDEKIRKOLEVIVSPSCQKQAGCVREITTK-----LGNEKOPTNP 596
Db 1090 DITVLGALLDTSKHFRLDKLCEKCAKCSIMIG-VLDVYKHEFKRTTYSENEVYDLND 1148
QY 597 FLEKIRKFL--ERIPVHIDTESIS--ALIKOVNKSIDGTADD-----EDEGVPT 642
Db 1149 SVQTIKELVWINDILVAPAFQSENPSCOLEVALVIOESLKYCGLSESMDMNHKELYPN 1208
QY 643 DOAIRAGLELLKVLFT--HPI--SFHSAETFEESLACKMDEKVAEA-----ALQ 690
Db 1209 EAKL--WEKFNVSQKTIITYPILSSLYLAOSWKETVP-LKPPSNFEEGQIYWKRTTLD 1264
QY 691 IFKNTG-----SKIEDEPPIRSALLPYLH--HKSCKGPP-----723
Db 1265 LK--TGTTENHPGHVFSLSLRREDGSLNFFLYSIDITIKAKGPPYADILNGIITEF 1323
QY 724 -----ROAKYAIHCHIAITSSKE-----TOFAQIETEPH-----KSLDSPN 759
Db 1324 DSIFTCNLEGMNNLQVDSLRCVSEISFIVEYCKRWATERKQNSKLHGFTIITDXTJN 1383
QY 760 LEHLITPLVITIGHITALDPOFAAPKSWATFIVKDLMLMDRLP-----804
Db 1384 M-----LIRIDEFLRTTPS-----DLIAQRSLETSPERSALYLRQCY 1421
QY 805 -----GKTTKLMVDEEV-----SPETVYKQIAIKMVRW-----835
Db 1422 RQNHDKNGOLKKNQIITYEEIGDIDSLGVLRTPATGNIVAKIELOQSEWKLAD 1481
QY 836 ---LLGKNNHSGSTLR-----LTTIHSDDGLTBOGKRS--KPDMSRLAAGSA 885
Db 1482 CFNVLGFSDDPKTTTMLKSMYDHQLYSQIISNSSFHSSDGISSLSPDVK-----1532
QY 886 IVKLAOPCYHEITTEBOYLALAINDE-CYQVROVFAQKLH--KGLSRLRLPLEY--- 939
Db 1533 -----EMYSIGLEANLEGVNQTLLKMWVQIESLRNIDREVLQYNIA 1576
QY 940 MAICALCAKDPVKEERRAHAQCLVKNINVREYIKQHAANSEKILSLPEYVVPYTHLL 999
Db 1577 KALIASNEBPLRTO-----KYIHSFRLGTGTFITSSKRTTLLKQNLMLKHS 1627
QY 1000 AHDPDYV-KYODIEJOLKDVKECLMEVLEIIMAKNENSHAF--IRKMTENIKOTKDAQSPD 1057
Db 1628 -YDLSFSSAKDFEYKSNMTTILDYRMRERIGADFVPNHYILISMKSPDQKMNQOA--- 1682
QY 1058 DAKMNEKLYVCDVANM-----IIMSKSTYSLE 1086
Db 1683 DADIGKTFPTLAQIARNNARLDIASESLMHCLE 1715

```

RESULT 11
 US-08-290-731C-2
 : Sequence 2, Application US/08290731C
 : Patent No. 5843646
 : GENERAL INFORMATION:
 : APPLICANT: BOWTELL, David Douglas Lawrence
 : TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE

```

: TITLE OF INVENTION: SON OF SEVENTEEN (MSOS) GENE,  

: TITLE OF INVENTION: AND MSOS POLYPEPTIDES  

: NUMBER OF SEQUENCES: 15  

: CORRESPONDENCE ADDRESS:  

: ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  

: STREET: 2100 PENNSYLVANIA AVENUE, N.W.  

: CITY: WASHINGTON  

: STATE: D.C.  

: COUNTRY: USA  

: ZIP: 20037  

: COMPUTER READABLE FORM:  

: MEDIUM TYPE: Floppy disk  

: COMPUTER: IBM PC compatible  

: OPERATING SYSTEM: PC-DOS/MS-DOS  

: SOFTWARE: PatentIn Release #1.0, Version #1.25  

: CURRENT APPLICATION DATA:  

: APPLICATION NUMBER: US/08/290,731C  

: FILING DATE: 17-OCT-1994  

: CLASSIFICATION: 435  

: PRIOR APPLICATION DATA:  

: APPLICATION NUMBER: PCT/AU93/00068  

: FILING DATE: 17-FEB-1993  

: PRIOR APPLICATION DATA:  

: APPLICATION NUMBER: PLO921/92  

: FILING DATE: 17-FEB-1992  

: ATTORNEY/AGENT INFORMATION:  

: NAME: KIT, Gordon  

: REGISTRATION NUMBER: 30,764  

: REFERENCE/DOCKET NUMBER: Q-36066  

: TELECOMMUNICATION INFORMATION:  

: TELEPHONE: (202) 293-7060  

: TELEFAX: (202) 293-7860  

: TELEX: 6491103  

: INFORMATION FOR SEQ ID NO: 2:  

: SEQUENCE CHARACTERISTICS:  

: LENGTH: 1319 amino acids  

: TYPE: amino acid  

: TOPOLOGY: linear  

: MOLECULE TYPE: protein  

: US-08-290-731C-2

```

```

Query Match 2.44; Score 176; DB 2; Length 1319;  

Best Local Similarity 19.24; Pred. No. 3.4e-05;  

Matches 252; Conservative 200; Mismatches 494; Indels 364; Gaps 68;  

QY 189 DTVSQELDLVIVLVNVAHNLKQA--YDLAALLKRTQAI-----EPYIT 225  

Db 174 DMHODVEDINILSLDEPSTSGEOTYDVKAFMAEIRQYIRELLIKVREPEVSN 233  

QY 236 FPNQVIMLGKTSISDSEHVFDDLILELYNIDSHLLSVLPOLFEKLSNDEERLQVVKL 295  

Db 234 -----SKLFSSNDVENFESRIVDIHELVSALLGHIEDTYEMTBEGSPH-----PL 278  

QY 296 LAKMFGAKDSELASQNKPLMQCYLGRPNIDHVRILECVFASHCL--MNHPLADLTE 353  

Db 279 VGSCEFEDLAEELA-----FDPEYSYARDILRP-----GFHGHFLSQSLSPGAAL---- 322  

QY 354 YLKVRSNDPEEARLHDVIVSYTAAKKDILLVNDH--LNFPRERTLDKMR--VKREAM 409  

Db 323 YLOSIGRGEFAOY-VLPRLILAPVYHCL--HYFELLIKOLEESSEDEDECKMQAI 377  

QY 410 MGLAQIYKRYKALQSAAGDAKDAKOIAMIKDKLHIYQNSIDRLAVE--RIFAQVAVPH 466  

Db 378 TALLNV-----QSGMKTSKSLA-----KRLSLSACRFYSQOMKKG 415  

QY 467 NLEETTERMKCLYYLYATLDLNAVKALEMMKCONLRLHQYKDLLD-----LIKOPTDAS 521  

Db 416 OL-----AIKKNNEIOK--NIDGWEEKDIOGCCNEFIIMGTILRVG 454  

QY 522 VK-----AIFSKYVAVITRN-----LPDPCKAQ-----DFMKKFTVOLEDE--KIRKQ 562  

Db 455 AKHERHIFLDGLMICKSNHGQRLPGASSAEYRLKEKFPMK--VOINDKDTSEKHA 513

```

QY	563	LEVV-----SPTSCQKQEGCVREITTKLGNPKQOTNPLEMI-----KELLERINAPVHD	614
Db	514	FEIILKQGNSTYIFSAKSAE-----EKNMMAALISIQYSTLERMLDVTVL	559
QY	615	TESISALIKQYKNSIDGTAD-DEDEGVATDOAI--RAGLELTK--VLESTPHISFHS--	666
Db	560	QEEKDEQNRPLSAEYVRAEPDSENNILFEENVQPKAPIITIKAGTVLKILERLYHNMTA	619
QY	667	---AEFF-----ESILACL---KMDKEVAEALQIFKNTGSKLIEDEPHIARS	708
Db	620	DPNVEFTEFLTYRSEFCRQOELLSTLIERFEIPEPERPEADRIAIGNDQPLSAELKRRPK	679
QY	709	ALLVLLHNHKKSKGGRPRQAKVAIHCHNAIFSSKEQFOAFQIFERLHSLDPSNLEHILTVLV	768
Db	680	EYI-----OPVOLRLVNCRHW----EHHFYD-FE---RDADLLQRMEEFIG	719
QY	769	TIGHITALIARPOFAAPMKSWAT--FYVKLLMDMRPLRGKTKTKLWAPDEEVSPTWVK	825
Db	720	TVRKRAM-----KKWVESITKIIQRRKKIARDNPGHNIT-----FOSSPT---	760
QY	826	IQAIKMYRWLLGKKNHNSKSG-TSTLRLLT-----TLIHSN-----GD	863
Db	761	-----VEMHI-----SRPGHLETFDLLTLHPRIEARQQLTLESDDLYRAVQPSLWGS	807
QY	864	L-TEGGKISLSDMSRLRLAGSALVYKLAQERPCYHEITTLDOYQCALAINDCQVQVQY	921
Db	808	VWKEDKEINSPN-LKMIKHITNLTLMFERKCIETVENLE---RAVAVSRIITELIQQ	861
QY	922	FAOKLHKGSLRLRLPLEYMAICALCAKDPV-----KERANAROCVLKNNINVARREYIK	974
Db	862	FOE-----LNNFNGVLE---VYSAMNSPVRRLDHTFQOISRQKKILEAHELSEDHVK	913
QY	975	QHAAVSEKLSSLREYVVPYTIHLL-----AHDDYVKVQDITQOLKDYKCSLWF-----	1023
Db	914	KYLA-----KLRSINPPCVPEFGIYLTILNLTKEGNEVLR-----RKGKELINRSKRR	963
QY	1024	VLEILMAKNENNSNAIFIKMKYENIKQTDAAQGPDDAKNNEKLYVCUVAAMINISKTTY	1088
Db	964	VAEITGTGEOQONOPYCLRVDPDIKRFEENLNPQNSM-EKEFT-----DYLEFKSLEI	1016
QY	1084	SLESKPQVLPAREFTQDPDKNFMTKNVLLPREMKSFFTPRGKPKTTNVLGAVNKRPL--SS	1140
Db	1017	EPRIKPP-LP-RT-----PKKISTYLPKSGVPARSNRP--GTYMHPRLIOE	1056
QY	1141	AGKQSQTSSSMETVSNANSSSSNPS--SPERIKGRLDSSSM-----DHSNEDYTMS	1191
Db	1060	PKRISYSRIPSEIESTVASANSRPTPLTPRPASGTSNTDVCSPVDDHDSAPRHSQA	1119
QY	1192	PLPCKSKDRDSDLVNSELKPRGKKTKVTEOEKIL--GMDDLTKVQEO-----K	1242
Db	1120	SVSSISLSKGTDEVPVPRPVPRRRRRESAPRESSPSKIMSKHLDSPRALPRQPTSKAYS	1179
QY	1243	PKGQSRKRRKHTTASEDSEQWPREKRLKEDLLENEDQNSPPKKGKRG-----	1291
Db	1180	PKRISIDTSTISDPRESRPLLRPREVATRPVVFSSPRLINLQRPRLGKSKDHGMAFPNSP	1239
QY	1292	---RPKPLGGGTPEKEEPTMK-----TSKSGSKKKSGPAPAREEHEEE	1331
Db	1240	SPFRPPR---QTPSPHGTNRHRLSPRLQEMDLHSLAGPVRVPRQSTSQ	1266
RESULT 12			
US-08-290-731C-6			
Sequence 6, Application US/08290731C			
Patent No. 5843646			
GENERAL INFORMATION:			
APPLICANT: BOWTELL, David Douglas Lawrence			
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE			
TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,			
TITLE OF INVENTION: AND MSOS POLYPEPTIDES			
NUMBER OF SEQUENCES: 15			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS			

STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentm Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PLO921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: Q-36066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1336 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-290-731C-6

Query Match	2.48;	Score 176;	DB 2;	Length 1336;
Best Local Similarity	19.28;	Pred. No. 3.4e-05;		
Matches	252;	Conservative 200;	Mismatches 494;	Indels 364;
			Gaps	68;

```
QY      189 DTYSGELDTVJLVNLYPAHKNLNKRA---YDLAKALKTQAQAI-----EPVIT 235
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      191 DMFHQVEINILISTLDEEPTSGGQTYYDVLKFMAMEIROYIELMLIKFVEPVS 250
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      236 FFRQVLMLGTSISDSIEHFPLILIELYNIDSHLLSVLPLEFKLSNDNEERLOAVKL 295
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      251 -----SKLFSSNDVENIFRSYVIDIHELSVKILLGHIEDPEVTMGDPH-----PL 295
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      296 LAMFGAKSSELASONKPLMOCYLCRFNDIHVPRLBECVKAHSCL--MNHPLADLTE 353
           |||||
Db      296 VGSCEFDLAEELA-----FDPEESYARDILRP-----GRHGFFLSQLSKPGALL---- 339
           |||||
QY      354 YLKRVRSDPEEAIRHDVIYSIVTAAKKRILLVNDH--LNFVREPRFLDKRMV-VRKEM 409
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      340 YLOSIEGRKEAVQY-VLRPLLAAVYHCL--HYFELLKQLEKSEDEDENCKMQAII 394
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      410 MGLAQYKYRYALQSAGDAKQAIWIKDKLLHYONSIDRLVE--RIFAQYVPH 466
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      395 TALLNV-----OSGMEEKICSILA-----KRLSESACRFSOOMKCR 432
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      467 NEEFTERMKCLYVATLDLVNVRKLVLENMKCONLIRVOYDILD--LIOPRTAS 521
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      433 QL-----AIKMNETQR-NIDGEGKDIOCCNEIMEGTLTRVG 471
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      522 VR----AIFSRYVNIYTRN-----LPDPGAQQ-----DEMKAFTOLEDEE-KIRQ 562
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      472 AKHERHIFLDFGLMICCSNHGOPTLPASSAEYRLKEKFPMR--VOINKDPTISEKHA 530
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      563 LEVLV-----SPSCOKOAGECVRELTTKTGPKOPTNPTEMI-----KELLEARIAPH 614
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      531 FEIILLDGNSVIFYSAKSAB-----EKNNMAALISLOYRSTLERMLDYVA 576
           |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      615 TESIALIKOVKSIDGTAD-DEDEGVPTDAI--RAGLELLK---VLSETHRISFSH-- 666
```

```
Db 577 OEEKEEQMLPSAEVYFAEEDSEENILFEENVOPKAGIPLIKAGTVLKLIERLTLYHMYA 636
Qy 667 -----AETP-----ESLACL-----KMDKAVADALQIKNTGSKIEEPPIHRS 708
Db 637 DPNFVFRTFTYTSFCRPPQELLSTLIERFEIPEPEEADRIALENDQOISALIKFRK 696
Qy 709 ALPLVLHKKSKKCPROAKYAIIHCIAIFSSKETQFOAIFEPPLIKSHSDPSNLHLITPLV 768
Db 697 EYI-----QPVOLRVLANCRHW-----EHMFVD-PE-----RDADLQJMEEFIG 736
Qy 769 TIGHIALLAPDQEAAPKKSVAAT--FIVKDLMDRLPGKKTTLKLVNPEEVSPEPMVK 825
Db 737 TVRGKAM-----KKWVESITKIIORKKIARDNGPGHNIT-----FOSSPPT--- 777
Qy 826 IQAKMMVRLGLGKNNHSG-ISTLRILT-----TILHSD-----GD 863
Db 778 -----VEMHI-----SRPGHETFDLTLHPLEIAROLITLESIDLVRAPVSELVGS 824
Qy 864 L--TEQGRISKPDMSRLRLAAGSAIVKLAOEPCEYHEITLLEQYOLCALAINDECYQVROY 921
Db 825 VMTFEDKEINSPLN--LKMIHRTNTLTFEKCIVEENLEE--RVAVSRITIELLOV 878
Qy 922 PAQKLHGLSLRLPLEYMAICALCAKDPV-----KERRAHROCLVKNINVRREYLK 974
Db 879 FOE-----LNNFNGVLE--VVSAMNSPYRLDHTFEQILPSROKILIEAHLELSEDDHYK 930
Qy 975 OHAAVSEKLSLDEYVVPTIHL-----AHDPDVYKVDIQLDVKECMLF----- 1023
Db 931 KYLA--KLNSINPCVPFGIYITLNLKTEEGNPEVL-----RHGKELINFSKRRR 980
Qy 1024 VLEILMAKNENNSHAFIRKMKWENIKOTKADQPPDAMKNEKLYTVCVAMNIIKSKSTTY 1083
Db 981 VAEITGELIQYQNGPYCLARVEPDILKRFENLNPNGSM-EKEFT-----DYLENNKLEI 1033
Qy 1084 SLESPKQPVLPARFTPOPKNFSTKNYLPPEMKSFPTPGKPKTTNVLGAVNKL---SS 1140
Db 1034 EPRHPRP--LP-RP-----PKKYSYPLKSPGVAPSPNRP-----GTWRHPTPLQOE 1076
Qy 1141 AGKOSOTKSSHMETVSNASSNSPS--SPGRIGRLDSSSEM-----DHSENDYDMS 1191
Db 1077 PRKISYRIPESEIESTASAPNSPRTPLTPPASGTSNTDYCVFSDSHASPFHRSKA 1136
Qy 1192 PLPGKSKDKRDDSDLVSELEKPRGKRTPYTQOEKL--GMDLTKLVOEQ-----K 1242
Db 1137 SVSISLSKGTDEVPPRPVPPRRRPSAPAESPSKIMSKHLDSPALIPRQTSKAYS 1196
Qy 1243 PKGQBRKRGHTASEDEQWPEEKRLKEDILENEDEONSPPKKGRG----- 1291
Db 1197 PRYSISDRTSISDPRESPLLPREPVRTPDVFSSPLHLQRPPLGKSKSDHGNAFFPNSP 1256
Qy 1292 ---RPPPLGGGTPEKEEPTMK-----TSKKGSKKKKSGPPAPREEESEE 1331
Db 1257 SPFTPPPP---QTSPSGHTRHRLPSPPLTQMDHLHSIAGRPVPRQSTISQ 1303
```

RESULT 13
US-08-973-462-8

Sequence 8, Application US/089734628

Patent No. 6191270

GENERAL INFORMATION:

APPLICANT: DRUTLHE, PIERRE

APPLICANT: DAUBERSTES, PIERRE

TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

FILE REFERENCE: 0660-0125-0 PCT

CURRENT APPLICATION NUMBER: US/08/973, 462B

CURRENT FILING DATE: 1998-02-06

EARLIER APPLICATION NUMBER: PCT/FR96/00894

EARLIER FILING DATE: 1996-06-12

EARLIER APPLICATION NUMBER: FR 95/07007

EARLIER FILING DATE: 1995-06-13

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 2.4%: Score 174; DB 3; Length 1786;
Best Local Similarity 17.0%; Pred. No. 8, 1e-05;
Matches 203; Conservative 192; Mismatches 398; Indels 402; Gaps 48;

```
Qy 101 DKLDIPM-----FITROLGLEDTSKPOFNRYFLLLENIA--WKSYNICELED 149
Db 838 EEIKDSILNEIEYKENVVTTILENVDETTAEVSFTPSNILEIEIENNTINDTIEEKL 897
Qy 150 SNEFTPOLYTPSVINNGHNOKYHIMVNDLMSIICEGTYSOELLDY-----LVN 202
Db 898 LHE-----NVLSALENTQSEEEKKEVDIVIEV--KEEVATTLIETVEAQAEKSANT 948
Qy 203 LVPAHKNLKOAYDLAKALKRTPQAIEPYITTFEFOVLMIGKTSISDLSEHVEDLITEL 262
Db 949 ITEIFENLEENAVESN--ENVAENLEKLENYETVTL-----DKVEET 989
Qy 263 YNIDSHLLSVLPOLFEKLSNDNEERLQVYVLLAKKAGKASELASQNKPLMQCYLGRF 322
Db 990 VEISGESLENNEKMAKFESEIFDNVKGIO-ENLLTGFRSIETSIYIQSEKVD-----L 1043
Qy 323 NDHIVPIRECVAFASCLMNHDP-----LAKDLELYKVSHPDEEAIIRDYIVSY 375
Db 1044 NENVVSSILNDIENEMKGLNKLLENISSTEGVQETVEH-----VEQNVYVDV- 1091
Qy 376 TAAKDLILVNDHLINVTRETDLRBVRVREKAMMGAJOYKYKVALQSAAGDAKQIAW 435
Db 1092 ---DYPAKDQPLGLINEA-----GLKEMFNLIEDYK----- 1122
Qy 436 IKDKLHIYYQNSIDRLIVERLIFAQYVPHNLETTERMKCLYLVATLDLNAVKALNEM 495
Db 1123 -----SESDVITVEIKEDE--PVQKE-----VEKEVYSIIEEM 1153
Qy 496 WKQCNLRHOVKDLDLIDKQPKTDASVKALFSKMYVITRNLPDGRKADQMKKFTOYLED 555
Db 1154 -----EENIVDLEEEKEDLDTKMDAAVESEIISDSKEETESIKDKDQVSLVVEE 1206
Qy 556 -----DEKIRKQLEVVSPQSCQOAEQCVRETKKLGNDKOPTNPF-LEMIKFLLEIRI 608
Db 1207 VQNDMDSEVKEVLELKNMEEELMKDAVE-INDTTSILIBETQOLNEVEADLK----- 1259
Qy 609 APVHIDYESISALIKOVNKSIDGTADDEDCGVPPTQAIRAGLELLKVLSTFHPISFSAE 668
Db 1260 ---DMEKLEKELEKALSEDSKETIIDAKD-----TLEKVIIEEHDT---T 1298
Qy 669 TFEISLACIMKMDKVAEAA-----IQIFKNTGSKIIEEDPHIRSLALPYLHN 716
Db 1299 TLDVEVELKQVEEKIEKVSOLKDLLEDILKEVKEILESEILEDYKELKTJETDILBE 1358
Qy 717 KSKKGPPOAKYAIHC-----IHAIFSSKETQFOAIFEPPLIKSHSDPSNLHLITPL 767
Db 1359 KKEIEKHFEKFEFEAEIEINDLEADILKESVLSLEEEKKLEVEHLEKE--EVEHIIIS-- 1414
Qy 768 VTIGHIALLAPDQ--AAPKSVAVATFIVKDLMDRLPGKKTTLKLVNPEEVSPEPMVK 825
Db 1415 -GDAHIGLEDEDDLEEVDDLGKSIDMLKGMDEIGDM--DKESL-----EDVYTKIGER 1465
Qy 826 IQAKMMVRLGLGKNNHSGSGTSLKLLTTLTILHSDDLTBQGRISKPDMSRLRLAAGSA 885
Db 1466 VESLKDVLSALGDEBQOMKTR----- 1487
Qy 886 IVKLAOECYHEITLLEQYOLCALAINDECYQVROVPAOKLHGSLRLPLEYMAICAL 945
Db 1488 ---KKAQRPKLEEVYLKE-----EVAEEKKKIITK--KAVRDIK----- 1522
Qy 946 CAKDPVERRAHAROCVKNINVRREYLKOHAAVSEKLSLDEYVVVPTIHLAHDPDY 1005
```

Db 1523 -DKEPKDE-----IYEVEMKDEDEED--VEED-----IEEDIEE 1554
 QY 1006 VKVDIOEQ--LKDYECIMFVLEILMAKNENNSHAFIRKMNENIKOTDAGCPDDAKANE 1063
 Db 1555 DKVEDIDEDIDEDIGEDKDEVIDLTVOK-----EKRIEKVAKK-----KLEK 1598
 QY 1064 K-----LYTVCDVAMNII-----MSKSTYSLESPKD----- 1090
 Db 1599 KVEEVSGLKHHVDVEMKYVOKIDKEVDEKVSKALESKNVDYTNVLKONODEFSKVNFK 1658
 QY 1091 -----PVLPA-----RFFP-----OPDKN----- 1104
 Db 1659 KYKVEAFPFISAVAFASYVGFEPFSLFSSCVTIASSSTYLLSKYDKIINKKERPFYSF 1718
 QY 1105 -----FSNTKNTLPPEMKSFPTPGKPKTTNVGAVNKKPLSSAGKOSQTSRMEYV 1155
 Db 1719 VEDIFKNLHYL--QOMKEKFS--KEKNNNVIEVTNKAERKGNVQVTNKTETKTKV 1770

RESULT 14
 US-08-588-985-2
 ; Sequence 2, Application US/08588985
 ; Patent No. 5777094
 ; GENERAL INFORMATION:
 ; APPLICANT: MICHIOYUKI MATSUDA et al.
 ; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Menderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/588,985
 ; FILING DATE: January 19, 1996
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1865 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL:
 ; ANTI-SENSE:
 ; FRAGMENT TYPE:
 ; ORIGINAL SOURCE: spleen cell of homo sapiens
 ; US-08-588-985-2

Query Match 2.4%; Score 172.5; DB 1; Length 1865;
 Best Local Similarity 17.6%; Pred. No. 0.00012;
 Matches 258; Conservative 212; Mismatches 441; Indels 553; Gaps 72;

QY 6 TRNDGKITPPGVKEISDKISKEEMVRLKMYVKTFFMDMDQDSPEEKELVTLNLHLAS 65

Db 803 TIVNDVKLVPRD--KELS-----KMFTEFLINVMGILFTICKVCLIEI--VHS 847
 QY 66 DPELKHGCKDVRLL-----LVACC--LADIPRI--YAPAPYTSPPDKLDI 106
 Db 848 DLFQHDRCREILLPYMTDQLKYLHERQDELEACCOLLSHILEVLYRKDVGPQO---RHV 903
 QY 107 FMFITROLKGLDEFPSPFNRYFYLEN--IAMVKSYNICFELEDSENEIFQTLYRTLPFSVI 165
 Db 904 QIMEKLLRTVNRVY--SMGRSELTGNFVACMTA--TLRQMEDYH--YAHLIKTE----- 954
 QY 166 NNGHNQYVMMHVD--LMSIICEGPTVQOE--LDPTVLNVLPAKHNKNOAYDAKALL 222
 Db 955 -----GKRRIVDVDELMTFTFMFKLIKKNYPRFPMVIMNVQ-----NKVFL 997
 QY 223 KRTAQAIEPYTTTFEYNOVLMGKTSISDLSEHVEDLILELYNIDSHLLSVLPQLEFKLK 282
 Db 998 R-----AINQYAD-----MLNKRFLDQAN--FE--LQLNWYFHLAVAEFTQESIQLE 1041
 QY 283 SINDNERLOVYKLLAKMGAKDSELASONKPLMOCYIGRFDNDIHYPIRLCEGVFASHCLM 342
 Db 1042 NFSSAKR--AKILNK--YGDNRROIGFEIRDMY--NLGQHK-----IKFL----- 1081
 QY 343 NHPDLAKDLTEYLKVRSHDPEEAIKHVY-----VSIVTAARKDILLVNDHL 389
 Db 1082 --PENAVGPILEMTLL-----PETELRKATIPPIFFDMQCEPSTRSFQMFENEIITKDHE 1135
 QY 390 LNFVREERTLDKRWVRKRAMMGLAQIYKVALQSAAGDAKQAIAMIKDKLLHYONSI 449
 Db 1136 VEGGRG--DEQYKVLFDKI--LLEHCRRHKYLAKTGTBFKLVY----- 1175
 QY 450 DDLRLVERIFPAQVWVRPHLETTERKCLYLYATLIDLNAVALNEMKCOYLRRQVYDL 509
 Db 1176 --RLMERLDDYRTIMDENKENRMSK-----TVNVLNFKYEIEREEMVIRLYK-L 1223
 QY 510 LDLIKQPTDASVKAIFSKVWVITRNLPDPKADQFMKFTVOLEDDKIKOLEVLVSP 569
 Db 1224 CDLHKE--CDNYTEAAYTLTL-----NAKLKMSDY----- 1253
 QY 570 TCSCKQAGCVREITKKLGNPKOPTN-----FLEMIKFLERIAVRNIDTESISALI 622
 Db 1254 -----CVANLQVRG--YQATQOQLKEQLYQELIHF----- 1284
 QY 623 KOVNSIDGTADDEGVPTQOATRAGIELLKVLSTFTPIFSHAEFTRESLACIKMDE 682
 Db 1285 -----DKGKMEBAIALKEL-----AEQYENM--FDYE 1312
 QY 683 KVAE--AAOIIFKNTGSKIE-----EDPHIRGALLPVLLHKKSGKGPPOAK 727
 Db 1313 QLSLELKKQAOFYENIYVIRPKRDYFVAVGYGQGFYFLAG--KVFIYRKEVEYPRFD- 1369
 QY 728 YAHCHIAHAFSSK--ETQFAQIFERLYHKLSDSNLEH-----LITPLVTIGHIALLA 777
 Db 1370 -----FEARLLTGFPRNEMKKMTTSPPDGDIKNSPGQYIQCFYVAPRKLD----- 1413
 QY 778 PDQFAAPKSNVAIF-----YKODLMDNRLPKKTKTLMVPRDEVSPEYTWKIQ 827
 Db 1414 PPKFHRPVSQIYVSFYRVNEVQREYSPIRKIGKENPDNEPANNMI--ERTIYTTAVKLP 1471
 QY 828 AIKMMVRYLGMKNHNSKSGTSLRLTLTIHSDGLTEQGIKSPDKSRLLAAGSAIV 887
 Db 1472 GI--LRRF-----EVSVMVEISPLENA----- 1493
 QY 888 KLAQPCYHEITLLEOYCALAINDECYQVQVFAOKLHGLSRLLRYLEMAICALCA 947
 Db 1494 -----LETQULTNDKIN-----SWVQOHLDDPSLPINPLMLNGI 1529
 QY 948 KDP--YKERRAHAROCLYKNINVRREYLYKONAAVSEKLLSLRPYVVPRTTILLAHDPYV 1006
 Db 1530 VDPAVVMGFAFYERKAFFTD-----RYLOEHEBAHEK----- 1560
 QY 1007 KVQDIEQDKVKECLM--FVLEITIMAKNENNSHA--FIKRVENIKOTDAGCPDDAK 1060

```

Db      1561 ----IEKLKDL--IAWQIFLAEGIRIHGDKVTEALRPPEHMEACFKOLKEKV----- 1608
QY      1061 MNEKLYTVCDAVMMNIIMSKSTIYSLSPKDPVLPAPFFTOPKNSNTNLYLPPEKKSFF 1120
Db      1609 ---EKELGV---RLTPSSLDROGRSPRSNV---RSFTMPSS----- 1641
QY      1121 TPGRKPTNVLGAVNKPLSSAGKQSQTKSSRMETYSNASSSSNPSSPGRIKGRLDSSEND 1180
Db      1642 -----SRLPSVA-----SVSSLSDSQSPRSRG----- 1663
QY      1181 HSENDYTMSSPLPGK---KSDKRDSDLVNSELKPPGR-----KTPYT 1223
Db      1664 ---SDFALEPLLPKMSRSODKLDKDLKEKKDKKKKSKHQEIEFEKEFKPTDIS 1720
QY      1224 -EOELGMDLTKIVQEQPK-----GSQSRKRGHTASRDEQWPEEKRLK----- 1271
Db      1721 LQOSEAVILSETISPLRPQKSVQVNMVIGSEKRSVSPSSSQQTPPPTPRALSS 1780
QY      1272 -----EDILENEDEONS-----PRKGRGRPPR 1296
Db      1781 MGSLELNGMTGADVADVPPLPLKGSVADYGNLMENODLIGSPPTPPPHQRLP-P 1839
QY      1297 LGGCTPKPEPTMTKSKGSKKSG 1320
Db      1840 LPSKTPPPPP-KTYRKQTSVDSG 1862

RESULT 15
US-08-971-988-2
; Sequence 2, Application US/08971988
; Patent No. 5786461
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,988
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/588,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1865 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: spleen cell of homo sapiens

```

```

US-08-971-988-2
Query Match      2.4% Score 172.5; DB 1; Length 1865;
Best Local Similarity 17.6% Pred. No. 0.00012;
Matches 258; Conservative 212; Mismatches 441; Indels 553; Gaps 72;

QY      6 TRINDGKITYPGVGEISDKISEMVRRLKAVNTFMDMDQSEEEKELYNLALHLAS 65
Db      803 TIVNDKYLVEDP--KELS-----KMETERILNVPGILLIQKLYCLIEL-VHS 847
QY      66 DEFLKHPGKDVKL-----LVACC--LADIFRI-YAREPRTSPDKADI 106
Db      848 DLFTQDCEKELILPMWTDQKYLHLEQEDLEACCOLSHILVILRKDGPQO---RHV 903
QY      107 FMEITROLKGLDTSQPFNRYEVLLEN-IAWVKNYICELEDSNEIFOTLRYLFSVI 165
Db      904 QIIMEKLYTNRTYI-SMGRSEILGNVACMTA--ILROMEDVN-YAHLIKTF----- 954
QY      166 NNGHNOKVIMHVD-LMSSIIGEDTVSQEL--LDFTVLNVLPAHKNLKOAYDLAKAL 222
Db      955 -----GKMRDQVVDLMEFTFIMEKNLIGKNVYRFDVIMNVQO-----NKVEL 997
QY      223 KRTQAQIEPYITTFNOVLMLGKTSISDSIENFDLLELYNIDSHLLSVLPQLEFKL 282
Db      998 R---AINQYAD-----MLKKKFLDOAN---FE--LDLMNMYELAVAFLTQESQLE 1041
QY      283 SNDNEERLOVYKLLAKMFGAKDSELASQKPLMOCYLGRFNDIHPVIRLECVKFSHCLM 342
Db      1042 NFSSAKR---AKILNK-YQDMRQIGFELRDMY--NLGQHK-----IKFI----- 1081
QY      343 NHPDLAKDLEYLKVRSHDEEAIKRDVI-----VSIYTAARKDILLVNDHL 389
Db      1082 --PEWVGLIEMTLI---PETELRKATIPRIFDMQCSFHNSTRSQFMENEITIKLDE 1135
QY      390 LNFVRERTIDKRRVYKEMAGLQYKKYVALDASAGKAAQOIAIKIKLHIYQNSI 449
Db      1136 VEGGRG--DEQKVLVDKI--LLECRNKKYLAKTGEFVYKLV----- 1175
QY      450 DORLVERIFQAQVYVHNLEETTERMKCLYLYATLDLNAVKALENMKCONLIRHQVKDL 509
Db      1176 ---RLMERLLDYTIMDENKEMRMC-----TVNVLNPFYKELEREMIRIYLYK-L 1223
QY      510 LDLIKQKTDASVKAIFSKVMVITRNLPDPAQODEFMKFTQVLEDEKIRQOLEVLVSP 569
Db      1224 CDLHKE--CDNYTEAAYTLIL-----HAKLLKMSEDV----- 1253
QY      570 TCSCKQAEQCVRIITKKNLPQPTNP-----FLEMIKFLERLAPVHIIDESTALI 622
Db      1254 -----CVAHLTQRDG--YQATTQOGLKEQLYQELIHYF----- 1284
QY      623 KOVNSIDGTADDEGVPFDQAIRAGLELLKVLSTPHPLSPHSAETPESLACLKMDDE 682
Db      1285 -----DKGMMEBALGKEL-----AQYENEM-----FDVE 1312
QY      683 KVAE--AALQIFKNTGSKIE-----EDFPHIRSALLPVLIHNSKKGPPRQAK 727
Db      1313 QLSLELKKQAQFYEIVLVKVRPDYFAVGYQGCPFTLRG--KVFIYRKGXEYERED- 1369
QY      728 YAIHCHIAIFSSK-ETQPAQIEPPLKSLDPSLHE-----LITPLVTIGHIALLA 777
Db      1370 -----FEARLLTQFPNAEKMKYTSPPGDINKSPQOYTQCTVTKRKLDD----- 1413
QY      778 PDQFAAPKMSWATF-----IVKDLMDRLPGKKTTLVAVPDEVSPEYWKIO 827
Db      1414 PKFHRPVSEQIVSEFRVNVQFESRPRKKEKPNDFANMMI--ERTYTTAVYKLP 1471
QY      828 AIRKAVRWLGMKNNSKSGSTLRLTLTILHSDGLTDEGKISKDMSRLRLAAGSAIV 887
Db      1472 GI---LRWF-----EVKSVFVEIISPLENA----- 1493
QY      888 KLAQEPCEYHEIITLBOYQALCALAINDECTQVROVYFQKLIKGLSRILPLEYATGALCA 947
Db      1494 -----IETMOLTNDKIN-----SMVOQHLDPSLPIPLMSLMLNGI 1529

```



```
QY 948 KDP-VKERRAHAROCQVKNINVRREYLKQHAAVSEKLSLLEPYVYPTIHLHLPDYV 1006
Db 1530 VDPVAMGSPANERKAFETD-----RYLOEHPAHEK----- 1560
QY 1007 KVODIEOLKDYKECIAW---FVLEILMAKNENNSHA---FIRKVENIKOTKDAQCPDDAK 1060
Db 1561 ----IEKIKDL-IAWQIPFLAEGIRINGDKYTEALRPFHERMEACFKOLKEV----- 1608
QY 1061 MNEKLYVCDVAMNITMSKSTYTSLESPPKDPVLPARFETQDPDKNFNTKNYLPPEMKSEF 1120
Db 1609 --EKEYGV-----RIMPSSLDDBRGSRPRSMV---RSFTMPSS----- 1641
QY 1121 TPGKRTTNVGAUVKPLSSAGKOSQTKSSRMEYTSNASSSSNPSSPGRIKGRLDSEMD 1180
Db 1642 -----SRPLSVA-----SVSSLSDSTPSRPG----- 1663
QY 1181 HSENEYTMSSPLPGK---KSDKRDSDLVRSLEKPRGR-----KTPVT 1223
Db 1664 ---SDGFALPEPLLPKMHRSQDKLDKDDLEKEKKDKKKKSKHQLFEKEFKPTDIS 1720
QY 1224 -E0EEKLGMDLTKLV0E0KPK-----GSQSRKRGHTASESD0QWPEEKRLK---- 1271
Db 1721 LQOSEAVILSETISPLRPQRPKSQVMNYIGSERRFSVSPSSQOTPPVTPRAKLSFS 1780
QY 1272 -----EDILENDEONS-----PPKKGKRGRRPKP 1296
Db 1781 MOSSLLENGMTGADVADYPPPLPLKGSVADYGNLMEODLIGSPTPPPPPHQRHLP-P 1839
QY 1297 LGGCTPKREPTMTKTSKSGSKKSG 1320
Db 1840 LPSKTPPPPP-KTTRKOTSYVDSG 1862
```

Search completed: September 24, 2003, 20:21:59
Job time : 35 secs

THIS PAGE BLANK (USPTO)

FT	Region	509..516
FT	/label= "Subdomain IV"	
FT	/note= "Hank's conserved region - beta strand 4"	
FT	Region	525..548
FT	/label= "Subdomain V"	
FT	/note= "Hank's conserved region"	
FT	Region	525..532
FT	/label= "Beta strand 5"	
FT	Region	540..548
FT	/label= "Al[ha helix D"	
FT	Region	554..567
FT	/label= "Subdomain Via"	
FT	/note= "Hank's conserved region - alpha helix E"	
XX		
PN	WO200050454-AI.	
PD	31-AUG-2000.	
XX		
PF	24-FEB-2000; 2000WO-US04732.	
XX		
PR	24-FEB-1999; 99US-0121461.	
XX		
PA	(TUFT) TUFTS COLLEGE.	
XX		
PI	Soto AM, Sonnenschein C, Geck P, Szelei J;	
DR	WPI; 2000-565451/52.	
XX	N-Psdb; AAA28051, AAA28052.	
PT	New human androgen-induced tumor suppressor cDNA sequence termed	
PS	'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient	
XX	diagnosis and treatment of prostate cancer -	
XX	Claim 2; Fig 1; 152pp; English.	
XX		
CC	This invention relates to a human androgen-induced tumour suppressor cDNA	
CC	sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located	
CC	on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell	
CC	proliferation and use as a marker for the efficient diagnosis and	
CC	treatment of prostate cancer. The invention includes AS3 cDNA and protein	
CC	sequences, a vector comprising the cDNA sequence, a host cell transfected	
CC	with the expression vector, and a method for producing an AS3 polypeptide	
CC	comprising culturing the transfected cells. AS3 has cyostatic activity,	
CC	and acts to suppress cell proliferation. The AS3 gene is useful as a	
CC	marker for the efficient diagnosis and treatment of prostate cancer. The	
CC	AS3 nucleic acid molecule can be used as a source of antisense agents for	
CC	sequence specific modulation of gene expression. The AS3 protein may be	
CC	used in the treatment of disorders caused by aberrant modification or	
CC	mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene	
CC	or aberrant post-translational modification of the AS3 protein. This	
CC	sequence represents the human AS3 protein sequence.	
XX		
XX	Sequence 1391 AA:	
SEQ		
	Query Match 100.0%; Score 7193; DB 21; Length 1391;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1391; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1 MAHSTRINDGKITTPPGVKETSDKISKEEMVRRLKNVTKTMDMDQOSEEKELYLNLA 60	
DB	1 MAHSTRINDGKITTPPGVKETSDKISKEEMVRRLKNVTKTMDMDQOSEEKELYLNLA 60	
OY	61 LHLADDFLKHGKVRLVLVACCLADIFRIYAPEAPYTSPDKIKDIFMFTBOLGLDEDT 120	
DB	61 LHLADDFLKHGKVRLVLVACCLADIFRIYAPEAPYTSPDKIKDIFMFTBOLGLDEDT 120	
OY	121 KSPQENRYFYLLENTIAWKYSYNICELEDSENEIFTQLYKTLPSVINNGHQVHHMVDL 180	
DB	121 KSPQENRYFYLLENTIAWKYSYNICELEDSENEIFTQLYKTLPSVINNGHQVHHMVDL 180	
OY	181 MSSITCEGDYQSOLLDTLVNLVVAHKNLNQAAYDLAKALLKRPAQAEPIITTFENVY 240	
DB	181 MSSITCEGDYQSOLLDTLVNLVVAHKNLNQAAYDLAKALLKRPAQAEPIITTFENVY 240	
OY	240 MSIIICEDDYQSOLLDTLVNLVVAHKNLNQAAYDLAKALLKRPAQAEPIITTFENVY 240	
DB	240 MSIIICEDDYQSOLLDTLVNLVVAHKNLNQAAYDLAKALLKRPAQAEPIITTFENVY 240	

QY	241	LMUKTISIDUSEVFDLIELYINIDSHLLSVLPÖEFKLSNDNEERLQVYKLLAKMF	300
Db	241	LMUKTISIDUSEVFDLIELYINIDSHLLSVLPÖEFKLSNDNEERLQVYKLLAKMF	300
QY	301	GAKOSELASONKPLMOCYIGREFNDIHPPIRECYKFPASHCLMNPDLAKDULETKYKRS	360
Db	301	GAKOSELASONKPLMOCYIGREFNDIHPPIRECYKFPASHCLMNPDLAKDULETKYKRS	360
QY	361	DPEEAIHRDVIYSIVTAAKDKILLVNDHLINFEVERPLDKRWVRKRAMMGLQIYYKYA	420
Db	361	DPEEAIHRDVIYSIVTAAKDKILLVNDHLINFEVERPLDKRWVRKRAMMGLQIYYKYA	420
QY	421	LQSAAGDAKQIAIMIKDKLHIYYONSIDRLLVERIIPAQIYWPANHLETTERMKCLYYL	480
Db	421	LQSAAGDAKQIAIMIKDKLHIYYONSIDRLLVERIIPAQIYWPANHLETTERMKCLYYL	480
QY	481	YATIDLIANVKLINEMMKONCLRHOVNDLIDLKOPRTDASVKAIFSKVWVITPNLDPG	540
Db	481	YATIDLIANVKLINEMMKONCLRHOVNDLIDLKOPRTDASVKAIFSKVWVITPNLDPG	540
QY	541	KAODFMKKFTQVLEBDEKIRKOLEVLYSPTCSCQAQAGVAREITKLGPNKOPTNPLEM	600
Db	541	KAODFMKKFTQVLEBDEKIRKOLEVLYSPTCSCQAQAGVAREITKLGPNKOPTNPLEM	600
QY	601	IKFLERIAIPHITOTESIALIKOYVNSIDSTADDEBQVPTDQAIAGLELLKVLVSFH	660
Db	601	IKFLERIAIPHITOTESIALIKOYVNSIDSTADDEBQVPTDQAIAGLELLKVLVSFH	660
QY	661	PISFSAETPESLACLKMDCKEVAEALQJFKNTGSKIEEDPHISALLPYLHHRSK	720
Db	661	PISFSAETPESLACLKMDCKEVAEALQJFKNTGSKIEEDPHISALLPYLHHRSK	720
QY	721	GPPROAKYAIHCIAIFSSKETQFOAQJIEPLHKSIDSNSNEHLITPVLVIGHIALLPDQ	780
Db	721	GPPROAKYAIHCIAIFSSKETQFOAQJIEPLHKSIDSNSNEHLITPVLVIGHIALLPDQ	780
QY	781	FAAPRKSWVAFFIYKDLMDNRLRGKKTTKLWVDEEVSPEPTWYKQAIKIMVWVWLLGKM	840
Db	781	FAAPRKSWVAFFIYKDLMDNRLRGKKTTKLWVDEEVSPEPTWYKQAIKIMVWVWLLGKM	840
QY	841	NNHSGSGSTRLLTTILHSDGDLTEOGKISKPRMSRLRLAAGSAIYKLAQEPCYHEIT	900
Db	841	NNHSGSGSTRLLTTILHSDGDLTEOGKISKPRMSRLRLAAGSAIYKLAQEPCYHEIT	900
QY	901	LEOYOLCALAINDCYOYROYFAOKLHKLSRLPLPYAICALCAKDPYKERRAHARQ	960
Db	901	LEOYOLCALAINDCYOYROYFAOKLHKLSRLPLPYAICALCAKDPYKERRAHARQ	960
QY	961	CLVKININRREYLRKOHAAVSEKLSLIPREYVAPTINHLADPRYVVOIDEOLKDVKEC	1020
Db	961	CLVKININRREYLRKOHAAVSEKLSLIPREYVAPTINHLADPRYVVOIDEOLKDVKEC	1020
QY	1021	LMFVLEILMAKNENNSHAFIRKMYENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIMSKS	1080
Db	1021	LMFVLEILMAKNENNSHAFIRKMYENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIMSKS	1080
QY	1081	TTYSLESKRDVLPARFPTQPDKNFSNKNLPLPEBMSFPTPGKPKTTNVLGAVNKRLLS	1140
Db	1081	TTYSLESKRDVLPARFPTQPDKNFSNKNLPLPEBMSFPTPGKPKTTNVLGAVNKRLLS	1140
QY	1141	AGKOSQTSRMEYVSNASSSNPSPRIKGRIDSSEMHSEMEDYTMSSPTLPGKKRSK	1200
Db	1141	AGKOSQTSRMEYVSNASSSNPSPRIKGRIDSSEMHSEMEDYTMSSPTLPGKKRSK	1200
QY	1201	RDDSDLVNSELKEPRGRKRTVTEQOEKLGMDLLTKLVQEQKPKQSORSRRKRGHTASESD	1260
Db	1201	RDDSDLVNSELKEPRGRKRTVTEQOEKLGMDLLTKLVQEQKPKQSORSRRKRGHTASESD	1260
QY	1261	EQOWPEERKLIEDLLENDEQONSPPKKGRGRPRPKPLGGGTPKKEEPTMKTSKGSKKSG	1320
Db	1261	EQOWPEERKLIEDLLENDEQONSPPKKGRGRPRPKPLGGGTPKKEEPTMKTSKGSKKSG	1320

Qy	1321	PPAAEEEEEEEROGSGNTFOKSKSQHVRYSRAAQRAESPESALIESITOSTFOKGRGRPSK	1360
Db	1321	PPAAEEEEEEEROGSGNTFOKSKSQHVRYSRAAQRAESPESALIESITOSTFOKGRGRPSK	1360
Oy	1381	TPSPSPORKNV	1391
Db	1381	TPSPSPORKNV	1391
RESULT 2			
ABG72803			
ID	ABG72803	standard; Protein: 1297 AA.	
AC	ABG72803;		
XX	24-FEB-2003	(first entry)	
DE	Human tumour suppressor SCC-112.		
XX	Human; SCC-112; cancer; apoptosis; allodynia; cell proliferation; degenerative disorder; metastasis inhibition; breast cancer; causalgia; kidney cancer; bladder cancer; pancreatic cancer; colon cancer; squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster; global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia; hypoxia-induced nerve cell damage; anxiety; diabetes mellitus; cardiac arrest; spinal cord lesion; lung cancer; tumour suppressor; neonatal distress; Alzheimer's disease; colon cancer; stomach cancer; multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome; Huntington's disease; Parkinson's disease; Korsakoff's syndrome; amyotrophic lateral sclerosis; cell survival; uterine cancer.		
XX	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	Domain	2..137	
FT	Domain	/label- RhogEF domain	
FT	Domain	127..651	
FT	Region	/label- N-adapctln_domain	
FT	Region	166..187	
FT	Domain	/label- Leucine_Zipper	
FT	Domain	249..350	
FT	Region	/label- SKP1_domain	
FT	Region	597..617	
FT	Region	/note- "PEST sequence site"	
FT	Region	858..865	
FT	Region	/note- "Tyrosine kinase phosphorylation site"	
FT	Region	920..926	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1030..1036	
FT	Region	/note- "Tyrosine kinase phosphorylation site"	
FT	Region	1143..1663	
FT	Region	/note- "PEST sequence site"	
FT	Region	1216..1277	
FT	Region	/note- "PEST sequence site"	
FT	Region	1225..1231	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1227..1230	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1228..1234	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1232..1235	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1251..1257	
FT	Region	/note- "Nuclear localisation signature sequence"	
FT	Region	1278..1297	
FT	Region	/label- Epitope_region	
FT	Region	/note- "Residues 1278 to 1297 are specifically claimed in claim 41"	
XX			
XX	WO200281641-A2.		
XX			
DD	17-OCT-2002.		

PF	08-APR-2002; 2002MO-US10850.
XX	
PR	06-APR-2001; 2001US-281780P.
PA	(GEOU) UNIV GEORGETOWN.
XX	
PI	Kasid UN, Kumar D, Ahmad I;
XX	
DR	WPI: 2003-103330/09.
XX	
XX	N-PSDB; AMX14056.
XX	
PT	New isolated SSC (undefined) tumor suppressor polypeptides and
PT	polynucleotides, useful for diagnosing, preventing or treating cancer
PT	or degenerative disease, e.g. Alzheimer's Disease, Huntington's
PT	disease, or multiple sclerosis -
XX	
XX	Claim 13; Fig 1B; 83pp; English.
XX	
CC	The invention relates to a new isolated polypeptide SCC-112. The SCC-112
CC	polypeptides and polynucleotides are useful for diagnosing, preventing or
CC	treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
CC	pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
CC	cancer, stomach cancer, colon cancer, lung cancer); or degenerative
CC	disease or disorder (e.g. global and focal ischemic and haemorrhagic
CC	stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
CC	damage, nerve cell damage caused by cardiac arrest or neonatal distress,
CC	epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
CC	pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
CC	hyperalgesia, allodynia, Alzheimer's disease, Huntington's disease,
CC	Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
CC	Korsakoff's syndrome). The polypeptides and polynucleotides are also
CC	useful for inducing apoptosis in cancer cells, increasing survival or
CC	proliferation of a cell, or inhibiting cancer cell proliferation and/or
CC	metastasis in a cancer patient. The polynucleotides can be used as probes
CC	to detect complementary nucleotide sequences, or as primers to obtain
CC	additional copies of the polynucleotides. SSC-112 may also be used for
CC	identifying drugs for treatment of cancers. The present sequence
CC	represents the amino acid sequence of human tumour suppressor SCC-112.
XX	
SO	Sequence 1297 AA;
Query Match	64.5%; Score 4642.5; DB 24; Length 1297;
Best Local Similarity	68.8%; Pred. No. 3; 7e-299;
Matches 905; Conservative 159; Mismatches 203; Indels 49; Gaps 11	
QY	31 MYRRLKMYVKTFMDMDODSEEEKLYLMLALHLADFLKNGKRYRLVYACCLADIRI 90
DB	1 MIKRLKMYVKTFMDMDODSEEEKOYRLALHLASEFLRNPNKQVRLVYACCLADIRI 60
QY	91 YAPEAPYSPKLCIDIFMETRQLGLGDDTSPQNRKYFYLENTAMVKSNIPELEDS 150
DB	61 YAPEAPYSHDKLDIFLPIRQLGLGDDTSPQNRKYFYLENTAMVKSNIPELEDC 120
QY	151 NEIFTOALRTLFVSINNGHOKVHHMAYDLMSIICEEDPTVSOELDLTVLNVVAHKNTL 210
DB	121 NEIFTOALRTLFVSINNGHOKVHHMAYDLMSIIMECGVYQELGLSTLNLGRAHKNTL 180
QY	211 NKQAVDLAKALLKRTAQALIEPYITTFENQVLMGLGTSISDLSHFVLDLILEYNTSD 270
DB	181 NKQSFDLAKVLLKRTVQRTIEACIANFENQVLYLGRSSVSDLSGEVFDLIOELFAIDPHIL 240
QY	271 LSVLPOLFEKLSNDNEERLQVVKLALMPGAKOSELSONKPKMQCYLGRNDIHVPR 330
DB	241 LSVMPOLFEKLSNDNEERLQVVKLALFSGKSDLAQNPKMQCYLGRNDIHVPR 300
QY	331 LECVAFASHCLMNHDPDLAKDLTEYLKVSHPDEEALIRHDVIVSIYTAKKDILVNDHLL 390
DB	301 LSVKAFASHCLMNHDPDLAKDLTEYLKVSHPDEEALIRHDVIVSIYTAKKDILVNDHLL 360
QY	391 NFVBERLDRKRWYRKEMGLAQIYKKYALQSAAGKQAKOIAMIKRKLHIYQNSID 450
DB	361 GFVBERLDRKRWYRKEMGLAQIYKKYCLHGEAGKRAAEKVSIMDKLHIYQNSID 420

Db	121	LMTIAKNLPDEPKADQFVKKFNQVLCDDDEKRLSQLELLISPCSCQADICVREIARKLA	180	KW	actinic keratosis; arteriosclerosis; atherosclerosis; buritis;
Qy	589	NKQPTNPFLMKIKFLIERIAVYHIDTESIALIKOVNKSIDGTADDEEGVPTDOAIRA	648	KW	clitrorrhosis; hepatitis; mixed connective tissue disease; MCTD;
Db	181	NKQPTNPFLMKIKFLIERIAVYHIDTESIALIKOVNKSIDGTADDEEGVPTDOAIRA	240	KW	myofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
Qy	649	GLELLKVSFTPIPSHAEFESLILACMKMDEKVAEALIOFRKTKGKIEEDPHIRS	708	KW	autoimmune disease; inflammation; acquired immunodeficiency syndrome;
Db	241	GLELLKVSFTPIPSHAEFESLILACMKMDEKVAEALIOFRNTGHKLTETDLPQIRS	300	KW	AIDS; Addison's disease; adult respiratory distress syndrome;
Qy	709	ALLPVLHNSKKGPPOAKYAIHCHIAIFSSKETQFAQTEFDLHSLDPSNLEHLLTPV	768	KW	allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
Db	301	TLIPILHOKAKRGTHQAKQAVHCHIAIFSNKEVQAOFFEDLSRLANDVPEQLITPLV	360	OS	Werner syndrome, trauma; human.
Qy	769	TIGHIALAPDOFAPAPKMSWATFYVKDLMDRLPGKTKTLWVPDEEVSEPTMYKIOA	828	XX	Homo sapiens.
Db	361	SLGHSMLAPDOFAPAPKMSWATFYVKDLMDNSTGKNGKNSPDEEVSEVYAKQA	420	XX	
Qy	829	IKMAYRWLLGMKNNSKSGTSLRLTLTILHSDGDLTEQKISKPMKSLRLAAGSALYK	888	XX	
Db	421	IKLVLRLMLGMKNNSKSGTSLRLTLTILHSDGDLTEQKISKPMKSLRLAAGSALYK	480	XX	
Qy	889	LAQEPCHYHITLLEOYOLCALAINDECYQVRFQAKLHGLSRLPLEYMAICALCAK	948	XX	
Db	481	LAQEPCHYHITLLEOYOLCALAINDECYQVRFQAKLHGLSRLPLEYMAICALCAK	540	XX	
Qy	949	DPVKERRAHAROCYLVKNINVRREYIKOHANVSEKLLSLPEYVPTTHLADHPDYKV	1008	XX	
Db	541	DPVKERRAHAROCYLVKNINVRREYIKOHANVSEKLLSLPEYVPTTHLADHPDYKRS	600	XX	
Qy	1009	QDIEOLKDYKCELMFVLETLMAKNNSHAFIRKKNVENTKOTKAOGPDADAKNEXLYTV	1068	XX	
Db	601	QDVODLRIKCELMFVLETLMAKNNSHAFIRKKNVENTKOTKAOGPDADAKNEXLYTV	660	XX	
Qy	1069	CDVANNIIMSKSTYSLSPKDPVLPAREFTOPDKNSNTKYNLPPMKSFPTPKPKPTT	1128	XX	
Db	661	CDVALCVINSKALCNADSPKDPVLPAREFTOPDKNSNTKYNLPPMKSFPTPKPKPRA	720	XX	
Qy	1129	NVLGVNKPPLSAGQOSQTKSSRMETVS--NASSSSNPSSGRIKGR-LDSSEMDHSENE	1185	XX	
Db	721	CVLGVNKPPLSAGQOSQTKSSRMETVS--NASSSSNPSSGRIKGR-LDSSEMDHSENE	780	XX	
Qy	1186	D--YTMSSPLPGKSKDKRDSLDLVSELEKPRGRKRTPTVQTEKLGMDLTKIVQEQKP	1243	XX	
Db	781	ENPVRIISVTPVKNID-----PYKNKE--INSDOAT-----QGN	812	XX	
Qy	1244	KGSORSKRGHTASESDE--QOMPEEKRLKEDILENEDQNSP--PKKGRGPPKPLG--CG	1300	XX	
Db	813	ISSORGRKRTVTAAGAENIQOKTDEK-----VDSGPPAPSKPRGRPPKSESQGN	863	XX	
Qy	1301	TPKEPTMKTSKSGSKSGPPAPEEEROSQNTG	1338	XX	
Db	864	ATKNDLMLKPKPKGRK---AAVGQESPGGLEAGNAK	897	XX	
RESULT 4					
AAB01382 standard; Protein: 854 AA.					
AAB01382;					
20-OCT-2000 (first entry)					
Neuron-associated protein.					
Neuron associated protein; NEUP; neurological disorder; epilepsy;					
ischemic cerebrovascular disease; stroke; cerebral neoplasm;					
Alzheimer's disease; Pick's disease; Huntington's disease;					
dementia; Parkinson's disease; demyelinating disease; meningitis;					
prion disease; Kuru; Creutzfeldt-Jakob disease; neurofibromatosis;					
cerebral palsy; muscular dystrophy; central nervous system; CNS;					
peripheral nervous system; PNS; myopathy; schizophrenia;					

FT Modified-site 697 /note= "Potential glycosylation site"
FT Modified-site 712 /note= "Potential phosphorylation site"
FT Modified-site 747 /note= "Potential glycosylation site"
FT Modified-site 750 /note= "Potential phosphorylation site"
FT Modified-site 757 /note= "Potential phosphorylation site"
FT Modified-site 798 /note= "Potential phosphorylation site"
FT Modified-site 800 /note= "Potential glycosylation site"
FT /note= "Potential phosphorylation site"
XX WO20003447-A2.
XX 15-JUN-2000.
XX 10-DEC-1999; 99WO-US30408.
XX 11-DEC-1998; 98US-0210083.
PR 11-DEC-1998; 98US-9123456.
PR 09-FEB-1999; 99US-0119365.
PR 16-MAR-1999; 99US-0124687.
XX (INCY-) INCYTE PHARM INC.
PI Tang YT, Yue H, Baughn MR, Hallman JL, Lal P, Au-young J, Yang J;
PI Lu DM, Azimzal Y;
XX WPI: 2000-423423/36.
DR N-PSDB: AAA47423.
XX New human neuron-associated proteins and polynucleotides encoding them,
PT useful for diagnosis, treatment and prevention of cell proliferative
PT disorders including cancer, neuronal and neurological disorders
XX Claim 1; Page 103-105; 1455p; English.
XX Human neuron-associated proteins (NEUAP) can be used for for
CC treating or preventing a disorder associated with decreased
CC expression or activity of NEUAP. Antagonists of NEUAP are useful for
CC treating or preventing disorder associated with increased expression
CC or activity of NEUAP. NEUAP or their fragments or derivatives are
CC useful for treating neurological disorder such as epilepsy, ischemic
CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC disease, Pick's disease, Huntington's disease, dementia and
CC Parkinson's disease. NEUAPs are also useful for treating other
CC demyelinating diseases, bacterial and viral meningitis, prion
CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC metabolic diseases of the nervous system, neurofibromatosis, other
CC developmental disorders of the central nervous system, cerebral
CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC other neuromuscular disorders, peripheral nervous system disorders,
CC inherited, metabolic, endocrine, and toxic myopathies, mental
CC disorders including mood, anxiety and schizophrenic disorders, a cell
CC proliferative disorder such as actinic keratosis, arteriosclerosis,
CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC hemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC disease, adult respiratory distress syndrome, allergies, ankylosing
CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC complications of cancer, hemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal, parasitic, protozoal, and helminthic
CC infections, and trauma. This protein was given the Incyte ID no.
XX 2888437CD1.
XX Sequence 854 AA;

Query Match 37.8%; Score 2715.5; DB 21; Length 854;
Best Local Similarity 62.2%; Pred. No. 1.7e-11;
Matches 543; Conservative 114; Mismatches 167; Indels 49; Gaps 11;
QY 474 MKCLYIYATLDNAVALNEMKCONLKHQKDLDDLKOPTDASVAIFSKVAVIT 533
DB 1 MKCLYLYASTLDPAVALNEMKCONLKHRELLDLKHOPTEANCSAMFGKMTIA 60
QY 534 RNLPPGKADPFMKFQVLEDEDEKIRKQLEVLVPTSCGKQAGCVREITKKGNKOP 593
DB 61 KNLPPGKADPFVKFQVLEDEDEKIRKQLEVLVPTSCGKQAGCVREITKKGNKOP 120
QY 594 TNPLEMIKELLERIAVVIDTESISALIKOVNSIDGADDEDEGVPTDQAIRAGLELL 653
DB 121 TNPLEMWKFLLEIRIADVIDSEASIALVLMKNSIGTADDEEGVPTAIRSGLELL 180
QY 654 KVLSTPPIPSHSETESLCLAKMDEKVAEALDIFKNYGSKIEDEPPHRSALLPV 713
DB 181 KVLSTPPTSHSETESLCLAKMDEKVAEALDIFKNYGSKIEDEPPHRSALLPV 240
QY 714 LHHKSKGPPROAKYAIHCHAIIPSKKEPOAFQFEEDLHKSILDSNLEHTLPTVTGHI 773
DB 241 LHKAKKGTPOAKOAVCHIAITNKEVLAQIFPELSLSINADVEQLITPLVSGHI 300
QY 774 ALLAPDQAPAPKSWAFTVTKDLMDRLPGKTKTLWVPDEEVSPTMYKIQAIKMY 833
DB 301 SMLAPDQAPAPKSWAFTVTKDLMDRLPGKTKTLWVPDEEVSPTMYKIQAIKMY 360
QY 834 RMLGKNNNSKSGTSLRLTLTHSDGLTEBGKISKPDMSRLRAASAIYKLAQEP 893
DB 361 RMLGKNNNSKSGTSLRLTLTHSDGLTEBGKISKPDMSRLRAASAIYKLAQEP 420
QY 894 CYHEITITLQALAINDECVQVROVFOKILKGLSRLLPLEYATGALCKADPYKE 953
DB 421 CYHEITITLQALAINDECVQVROVFOKILKGLSRLLPLEYATGALCKADPYKE 480
QY 954 RRAHAROCLYKNINVRREYLKQHAASEKLLSLPEVVPYTHLLAHDPYKVQDIEQ 1013
DB 481 RRAHAROCCLKNISIRREYIKQNPMAFEKLLSLPEVVPYTHLLAHDPYKVQDIEQ 540
QY 1014 LKDYKECLMYLELTLAKNENSHAFTRKVENIKQTKRDQAGPDAAKNEKLYVCYAM 1073
DB 541 LRDKEELIMLELTLAKNENSHAFTRKVENIKQTKRDQAGPDAAKNEKLYVCYAM 600
QY 1074 NIIMSKSTYSLESPKQVLPAPFETQPDKNFSTKRYLDPPEMSFTTPGKPTTVIGA 1133
DB 601 CVIMSKSALCNADSPKQVLPAPFETQPDKNFSTKRYLDPPEMSFTTPGKPTTVIGA 660
QY 1134 VNKPLSSAGQSQTKSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENE--YT 1188
DB 661 VNKPLSSAGQSQTKSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENE--YT 1240
QY 1189 MSSFLPKKSKRSDSLVSELEKPRGKRTPTTBEDEKLGMDLTKLYQEQPKRSOR 1248
DB 721 IISVTPVKNDI-----PKYKNE--INSQAA-----QGNISSDR 752
QY 1249 SRKRGHVASDSE--QOMPEEKRLKEDILENDEBONS--PKKGRGPRPKPLG--GTPKEE 1305
DB 753 GKRTTYAAGAEINQQTDEK-----VDESGPAPAPSKPRGRKPKSESGQNAIKND 803
QY 1306 PTMKTSSKSKSGKPPAPEEEEROSGNTGTE 1338
DB 804 DLNKPINKGRRR---AAVQESPGLEAGNAK 832
RESULT 5
ABB67869
ID ABB67869 standard; Protein; 1218 AA.
XX ABB67869;
AC
XX 26-MAR-2002 (first entry)
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 30399.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 XX
 DR N-PSDB; ABL11972.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure: SEQ ID NO 30399; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AB57737-AB572072).
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 1218 AA;
 SQ
 Query Match 27.6%; Score 1982.5; DB 22; Length 1218;
 Best Local Similarity 37.3%; Pred. No. 1,3e-122;
 Matches 444; Conservative 247; Mismatches 421; Indels 77; Gaps 22;

DB 362 VVAIVETAKRDFTVLEAPDLEIVRETLRDLKKYIRRDANGLAIYKRAICEPNDLST 421
 QY 429 AAK-OIAWTKDKLHIYTONSIDRLYERIPAOYVPHNLETTBRMKCLYLYVFTDLN 487
 DB 422 GLKRVSDWIKNKLHGCVKGLVEDRLVERLLITGLVPKLAPERRMKLYHLGLDLDAN 481
 QY 488 AVKALNEMKKCONLRLHGVKDLDDLIKQPK-TDASVKAFIFSVMVITRNLPPGKAQDFM 546
 DB 482 ATKAFVELOKQMKTRNTVSDWIKLHHSKEFTPRVLSAQAANIATKLDPDLAAEYL 541
 QY 547 KKFTQVLEDEKERIRKOLEVLVSPSCQAQSCVREITKKLGPNKOPTNPELEMTKFFLE 606
 DB 542 TQFSNNLRKRDQGLKLCINIVLKRQVSCRECAQPMGLLKGAAHQ-SLYNYNTYKMLE 600
 QY 607 RIAPVHIDTESIALIKOVNKSIDGTADDEDCVPTDQAIKAGLELLKVLSTPHRISHS 666
 DB 601 RVASVAVPKESIGVILSLIEQCIKESGMEIGISAQEGERGELKMLSVFSANHFPT 660
 QY 667 AETFEILLACLRKMDKEVAEALOLFKNMG---SKIEEDFPHIRSAALLPVLHKRSKGGP 723
 DB 661 DTSLRHLISLSYEDDYVAPLVKTLTHLGRQPLIDPTPAIDELAPVCKDFALLIGTP 720
 QY 724 RQAKYAIHCIH-----AIFSSKETQ-----FAQIFEPHLKSLDPSNLBHL 763
 DB 721 KOAKHAVKCIFFVNSQSSASTDQATSGAGSASTPTQTVPHFINEIETRLKLTLP-NCEHQ 779
 QY 764 ITPLVITIGHIALAPDOFAAPKRSVAFIYKDLMDRLRCKTKTKL-----WPDEEV 818
 DB 780 RRTVITLGHIAVNMPOAFLTPKNNIARIYKELLIOE-VPAQORYELPEDSQCAQERL 838
 QY 819 SPEWTKIOATIKMWRVLLGKNNKSGSTRLRLTLTHSDGLTQEGKISKPDMSRL 878
 DB 839 PPDTICKIDALKAMARMLGLRTD-EHAQAQTFRLAFAVQNRGDLLONRKCGAKESWL 897
 QY 879 RLAAAGSAIVKLAQEPYHEIITLQYOLCALAINDECYQVQVFAQKLHKLGSRLR---- 934
 DB 898 RLGAACAMLKVCQGVGSDQYSAEQYLQLSQMAADPVEVEIRFARKLHKLGSRLPRNC 957
 QY 935 LPLEFMALCALCAPVVERRAHACQLYKNTINVRRETLKQHAANS-----EKLISLLP 988
 DB 958 LPDFMGLYVLAGLETERRKLDLVNRYVAETVQNVKRELYKTVAMTSPDSSTESQSLHILP 1017
 QY 989 EYVVPYTIHLHADDDYKVDIDQKQVKECLMVLLELMAKNNSHAFIRKVENIK 1048
 DB 1018 DYMLAFAIPVLVHPDRFNNHEDYVQLRKMEKCLRTLEPLAKRTTFYHSYKQLQILK 1077
 QY 1049 QTKDAQGPDDAKMNKELYTCVDANMIISK-----STYVLESPPDVLPAHFPTQP 1101
 DB 1078 HREFSLG-SDRKDNVKKMALCDLAMYIIDSKFSPPDGMTSFSMPL-----ALPEMYKEP 1132
 QY 1102 D-KNPSNKNLUPPEMKSFPTPGKPKTINVLGAVNKKPLSSAKQSOTS 1149
 DB 1133 AVANFQNDNVYIPLDVY-----LGA--KSTYSKAATAMTTS 1167
 RESULT 6
 AAU27827
 ID AAU27827 standard; Protein: 217 AA.
 AC AAU27827;
 DT 18-DEC-2001 (first entry)
 XX
 DE Human full-length polypeptide sequence #152.
 XX
 KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
 KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
 KW nervous system disorder; inflammatory disorder; cell differentiation;
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
 KW cytostatic; antirheumatic; antiarthritic; vulnery; antiinflammatory;
 KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;

KW neuroprotective; osteopathic; antidiabetic; antisthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX Homo sapiens.
XX NO200164834-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04926.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX 17-JUN-2000; 2000US-0597707.
XX 14-JUL-2000; 2000US-0616807.
XX 19-SEP-2000; 2000US-0664641.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Dmanac R;
XX
XX WPI; 2001-569862/66.
XX
XX N-PSDB; AAS44727.
XX
XX Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
XX Claim 10; SEQ ID NO 324; 153pp; English.
XX
XX Sequences AAU27676-AU28019 represent full-length polypeptides and
CC contig polypeptides of the invention. The proteins and their associated
CC DNA sequences are useful for the treatment, diagnosis and prevention of
CC various types of disorder in a mammalian subject such as a human, dog,
CC monkey, mouse, hamster or rat. The disorders include cancers such as
CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as hepatitis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 217 AA;
Query Match 11.7%; Score 839; DB 22; Length 217;
Best Local Similarity 100.0%; Pred. No. 9.7e-48;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
AAB58462
ID AAB58462 standard; Protein: 363 AA.
XX
XX AAB58462;
XX
XX 14-MAR-2001 (first entry)
XX
XX
DE Lung cancer associated polypeptide sequence SEQ ID 800.
XX
XX Human; lung cancer associated protein; neuroprotective; cytoskeletal;
KW cardiocactive; immunomodulatory; muscular active; vulnereary;
KW gastrointestinal; nephrotoxic; antinefective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
XX Homo sapiens.
XX
XX OS
XX
XX WO20005180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
PI Ruben SM;
XX
XX WPI; 2000-587514/55.
XX
XX N-PSDB; AAF18338.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX
PS Claim 11; Page 1337-1339; 1425pp; English.
XX
XX polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardiocactive;
CC immunomodulatory; muscular active general; vulnereary; gastrointestinal
CC general; nephrotoxic; antinefective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
XX
SQ Sequence 363 AA;
Query Match 10.5%; Score 758.5; DB 21; Length 363;
Best Local Similarity 45.0%; Pred. No. 4.3e-42;
Matches 172; Conservative 56; Mismatches 105; Indels 49; Gaps 11;

QY 1231 MDDLTKLVQOKPKGSGRSRRKRGHTASESDQOMPEEKRLKEDILENEDQNSPPKKGKR 1290
DB 1 MDDLTKLVQOKPKGSGRSRRKRGHTASESDQOMPEEKRLKEDILENEDQNSPPKKGKR 60
QY 1291 GRPPKPLGGGTPEEPKMTSKKSGKSGPPAPAEPEEEROSGNTQSKQOHVSR 1350
DB 61 GRPPKPLGGGTPEEPKMTSKKSGKSGPPAPAEPEEEROSGNTQSKSKQOHVSR 120
QY 1351 RAOORAEPSSESAIESTQSTPOKGRGRPKSTPSPQKKNV 1391
DB 121 RAOORAEPSSESAIESTQSTPOKGRGRPKSTPSPQKKNV 161

QY 965 NINVRREYLKQHAHVSEKLLSLPEYVVPPTIHLHADPDYVYVODIEQKDVKECLMEV 1024
DB 1 NISIRREYIKQNPMPATEKLLSLPEYVVPPTIHLHADPDFTSQDVQDRDIKECLMEV 60
QY 1025 LETIMAKENNSHAFTIKKVENIKQTKDAQGPDDAKNEKLYVGVAMNITMSKSTTVS 1084


```

XX      SQ      Sequence      333 AA;
Query Match
Best Local Similarity 42.7%; Pred. No. 2.6e-34;
Matches 150; Conservative 51; Mismatches 101; Indels 49; Gaps 11;

QY      996  IHLHADPDVVKVQDIEQLKDKVECLFVLEIMAKNENNSHAFIRKRVENIKOTKDAQG 1055
Db      2  IHLHADPDVTRSDQVDQLRDIKECLFMLEVLMTKNENNSHAFIRKRVENIKOTKDAQG 61

QY      1056  PDDAKNNEKLYTCVDAAMNIMSKSTYTSLESFQDVLPAFRTQDPKNSNTKNVLPPE 1115
Db      62  PDSEKNEKLYTCVDAVALCYNKSKALCNADSPKDPVLPKPFQPEKDCNCKSYTSEE 121

QY      1116  MKSEFIPGPKRTNIVGAVNKPUSAGKQSKQTSRMEYTS--NASSSSMPSGRKIKGR 1173
Db      122  TRVLLLTGKRPKPGVIGAVNKPUSATGRKPYRSTGTETGSNINVENLSPGKNSREQ 181

QY      1174  -LDSSEMDHSENEED--YTMSSPLPGKSKSDKRDDSLVRSLEKPRGKKTPTVEOEKLG 1230
Db      182  SSEAAETGVSENEENPVRIISVTPVKNID-----PVKNKE--IN 218

QY      1231  MDDLTLVDEQKRGKSGORSKRGHTASEDE--QQMPEEKRLKEDILENEDEQNSP--PKKG 1288
Db      219  SDQAT----QGNISDRGKRTVTAGAENIQOKTDEK-----VDESGPAPASKP 264

QY      1289  KRGRPPKPLG--GGTPKEEPTMTSKKSGKSKSGPPAPEEEREEEROSGNT 1338
Db      265  RRORRRKSESQGNATKNDLNRKINKGRK---AAVGQSPGGLPAGNAK 311

RESULT 10
AA001759
ID      AA001759 standard; Protein: 165 AA.
XX
XX      AA001759;
XX
XX      18-JUL-2001 (first entry)
XX
XX      Human secreted protein #38.
XX
XX      Human: secreted protein; immunogen; antibody; diagnosis;
XX      rheumatoid arthritis; hyperproliferative disorder; neoplasm;
XX      cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
XX      angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
XX      fungal infection; corneal infection; wound healing; cell culture;
XX      epithelial cell proliferation; skin ageing; transplantation;
XX      tissue regeneration; chemotaxis; food additive.
XX
XX      Homo sapiens.
XX
XX      WO200123546-A1.
XX
XX      05-APR-2001.
XX
XX      26-SEP-2000; 2000WO-US26323.
XX
XX      27-SEP-1999; 99US-0155805.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      PI      Rosen CA, Ruben S, Komatsu S, GA:
XX
XX      WPI: 2001-266150/27.
XX      DR      N-PSDB; AAS02396.
XX
XX      Nucleic acids encoding 37 human secreted polypeptides, useful for
XX      PT      preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX      PT      disease and diabetic retinopathy -
XX
XX      PS      Disclosure; Page 454; 494pp; English.

```

```

CC      CC      The sequence represents a human secreted protein of the invention. The
CC      CC      polynucleotides, polypeptides and antibodies raised against them are used
CC      CC      to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC      CC      rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC      CC      polynucleotides and antibodies are also used in diagnosing a pathological
CC      CC      condition or susceptibility to a pathological condition. The antibodies
CC      CC      can also be used in alleviating symptoms associated with the disorders
CC      CC      and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC      CC      immunoassay assays (ELISA). Disorders which are diagnosed or treated
CC      CC      include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC      CC      disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC      CC      e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC      CC      angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC      CC      infections caused by bacteria, viruses and fungi and ocular disorders
CC      CC      e.g. corneal infection. The polypeptides can also be used to aid wound
CC      CC      healing and epithelial cell proliferation, to prevent skin aging due to
CC      CC      sunburn, to maintain organs before transplantation, for supporting cell
CC      CC      culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC      CC      polypeptides can also be used as a food additive or preservative to
CC      CC      increase or decrease storage capabilities.
XX
XX      SQ      Sequence      165 AA;
Query Match
Best Local Similarity 65.0%; Pred. No. 5.7e-32;
Matches 106; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY      964  KNINVRREYIKQAAVSEKILSLPEYVYVYTHLLAHDPDYKQVODIEQLKVKKCLWF 1023
Db      1  ENISIRREYIKQPMATEKLSLPEYVYVYTHLLAHDPDYKQVODIEQLKVKKCLWF 60

QY      1024  VLEILAKNENNSHAFIRKRVENIKOTKDAQGPDPAKMKNEKLYTCVDAAMNIMSKSTY 1083
Db      61  MLEVLMTKNENNSHAFIRKRVENIKOTKDAQGPDPAKMKNEKLYTCVDAAMNIMSKSTY 120

QY      1084  SLESPKDPVLPARFTQDPKNSNTKNVLPPEKKSFTTPGKPK 1126
Db      121  NADSPKDPVLPKMFQPEKDFCNDKSYISETRVLLLGKPK 163

RESULT 11
ABB89260
ID      ABB89260 standard; Protein: 147 AA.
XX
XX      ABB89260;
XX
XX      24-MAY-2002 (first entry)
XX
XX      Human polypeptide SEQ ID NO 1636.
XX
XX      Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX      antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX      vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX      cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX      neurological disease; infection; human; secreted protein.
XX
XX      Homo sapiens.
XX
XX      WO200190304-A2.
XX
XX      29-NOV-2001.
XX
XX      18-MAY-2001; 2001WO-US16450.
XX
XX      19-MAY-2000; 2000US-205515P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      PI      Birse CE, Rosen CA;
XX
XX      WPI: 2002-122018/16.
XX      DR      N-PSDB; ABB89669.
XX

```

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -

PS Claim 11: SEQ ID NO 1636; 2081bp + Sequence Listing; English.

XX
XX
CC The invention relates to novel genes (ABU9449-ABU90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 147 AA:

Query Match 8.2%; Score 591; DB 23; Length 147;
Best Local Similarity 96.6%; Pred. No. 1.6e-31;
Matches 115; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MAHSKTRNDGKITPPGVKEISDKISKEEMVRLKMWVKFMDMDQSESEKELYLMLA 60
Db 29 MAHSKTRNDGKITPPGVKEISDKISKEEMVRLKMWVKFMDMDQSESEKELYLMLA 88
Oy 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKDIFETIROLKGLD 119
Db 89 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSPDKLKDIFETIROLKGLD 147

RESULT 12

AA65247
ID AAY65247 standard; Protein: 101 AA.

XX
AC AAY65247;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1408.

XX
KW Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.

XX OS Homo sapiens.

PN W09953051-A2.

PD 21-OCT-1999.

PE 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-038446/03.

DR N-PSDB: AA242861.

XX
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
XX
PS Claim 3; Page 776; 837pp; English.

CC AA24265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AA165438 represent the EST-related proteins corresponding to AA24265 to
CC AA243075. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA24249 to AA24264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 101 AA:

Query Match 7.1%; Score 513; DB 21; Length 101;
Best Local Similarity 99.0%; Pred. No. 1.5e-26;
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MAHSKTRNDGKITPPGVKEISDKISKEEMVRLKMWVKFMDMDQSESEKELYLMLA 60
Db 1 MAHSKTRNDGKITPPGVKEISDKISKEEMVRLKMWVKFMDMDQSESEKELYLMLA 60
Oy 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSP 100
Db 61 LHLASDFLKHGKDVRLVACCLADIFRIYAPAPYSP 100

RESULT 13

AAG02811
ID AAG02811 standard; Protein: 92 AA.

XX
AC AAG02811;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 6892.

XX
KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PE 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR N-PSDB: AAC02817.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13: SEQ ID 6892; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

SQ Sequence 92 AA;

Query Match 6.5%; Score 467; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.5e-23;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 839 MKNNHSGSTSTLLTLTILHSDGDLTEQKISKPMRSLRLAAGSAIVKLAQEPCHET 898
DB 1 MKNNHSGSTSTLLTLTILHSDGDLTEQKISKPMRSLRLAAGSAIVKLAQEPCHET 60
OY 899 ITLEQYQCALAINDECYQROYFAQKLHKG 929
DB 61 ITLEQYQCALAINDECYQROYFAQKLHKG 91

RESULT 14
AAU27999
ID AAU27999 standard; Protein; 111 AA.

AC AAU27999;
XX
DT 18-DEC-2001 (first entry)
XX

DE Human conitig polypeptide sequence #152.
XX
XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytoskeletal; antirheumatic; antiallergic; vulnery; antinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteoplastic; antidiabetic; antialstematic; antiallergic;
KW immunostimulant; analgesic; gene therapy.

XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-0504926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI: 2001-589862/66.
DR N-PDB; AAS44899.
XX
PS Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -

XX Claim 10; Page 144; 153pp; English.

CC Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC contig polypeptides of the invention. The proteins and their associated
CC DNA sequences are useful for the treatment, diagnosis and prevention of
CC various types of disorder in a mammalian subject such as a human, dog,
CC monkey, mouse, hamster or rat. The disorders include cancers such as
CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC bowel disease. The sequences exhibit activity relating to angiogenesis,
CC cell proliferation, cell differentiation, stem cell growth factor,
CC activin or inhibin. Therefore, they can be used to manipulate stem cells
CC in culture to give rise to neuroepithelial cells that can be used to
CC augment or replace cells damaged by illness, accidental damage or genetic
CC disorders. The sequences may also be used for regeneration of bone,
CC cartilage, tendons and ligaments and in tissue repair and burn healing.
CC Note: Some sequences for this patent did not form part of the printed
CC specification, but were obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 111 AA;

Query Match 3.8%; Score 276; DB 22; Length 111;
Best Local Similarity 100.0%; Pred. No. 9.1e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1337 TEOKSKSKOHVRRAQORAEPSSESAIESTOSTPOKGRKRPSTPSPKKNV 1391
DB 1 TEOKSKSKOHVRRAQORAEPSSESAIESTOSTPOKGRKRPSTPSPKKNV 55

RESULT 15
AAB95449
ID AAB95449 standard; Protein; 233 AA.

AC AAB95449;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:17905.
DE Human protein sequence SEQ ID NO:17905.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX

THIS PAGE BLANK (USPTO)

MEDLINE PUBMED REFERENCE AUTHORS	20499374 11042159 3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL PUBMED REFERENCE AUTHORS	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, U., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyoh-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.
TITLE		Functional annotation of a full-length mouse cDNA collection
JOURNAL PUBMED REFERENCE AUTHORS	Nature 409 (6821), 685-690 (2001) 21085660 11217851 5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL PUBMED REFERENCE AUTHORS	Nature 420, 563-573 (2002) 6 (bases 1 to 3313)	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaki-Ashira, S., Takekoshi, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (16-JUL-2001) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT		CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
		Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hnako Atsugi City,

FEATURES	source	Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers
	1. .3313	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="FANTOM:DB:B130042B12"
		/db_xref="taxon:10090"
		/clone="B130042B12"
		/tissue_type="parthenogenote"
		/clone_1lb="RIKEN full-length
		/dev_stage="9.5 days embryo"
		<1. .2476
		/note="unnamed protein product; 49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homo [Homo sapiens] (SPR1096KV4, evidence: FASTV, 92.3%id, 100%length, match=1430) putative"
		/codon_start=2
		/protein_id="BAC32242.1"
		/db_xref="GI:26337115"
		/translation="VKSIDGTADDEDEGVPTQAIKAGELKLSFTPIPSHSAE TPFSLLAKENDEKVAEALQIFKNTGSEIEDPPIKRSALPLHKRSGKPPRA KVAIHCHAFESSEKTOFQIPEPLKSLDPSLSEHLIPLYIGHALLAPOPAP LKSLVATFIYKDLMDRLDPLGKTTKLYNVDSEVSEPTWIKQAIKMYRWLGMKN HSGVSTLELTLTTHSDDLLEQGTSLKSPDLSKRLAGSAIVKADQPCHEIIT LEOYQOLCALINDECYQVRFQVFKLHKGSRRLPLEVATICALAKDPKERRRA HRCOLVNIYTRRYLKOHAASEKLSLEPYVPTIHLADPVPYKVDIEQLDK VKCLNFEVLEILAAKNENSHAFIRKVENIKOTKACQPDYKMEKLTVCDAVN IISKSTYSLESPKDPVLPAREFTQKNSNTKYLIPPEKMSFTPGKPKXTANYS AVNPKPLSAAKQKOTSSRMETYSNASSSSNPSRRIRGRDSSSEMDSENDYMS SPLPKKSDREDPDLVRSLEKPRSKKAPVTDPEKLSMDLITLVQDKGKSOR GKRGRGASDSDQWPEEKRRHEELLENDQNSPPKGRGRPRPKPLGGTSKEE TMTKTSKGNKKRLVPPVDDDEEERQIGNTERTSKSKQHTSKRAQRAESPESTAV ESTQSTPQKGRGRSKAPSPSQPKIRVGRSKVATKENDSSSEMDVLAQSSPVD TPOEGAEEDISGVNRRRSKRERR"
		polyA_signal 3288..3293
		polyA_site 3313
		/note="putative"
		/note="putative"
		BASE COUNT 1066 a 709 c 732 g 806 t
		ORIGIN
		Query Match 45.5%; Score 1899.6; DB 11; Length 3313;
		Best Local Similarity 89.2%; Pred. No. 0;
		Matches 2049; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
QY	1872	AGTGAACCAATCAATAGATGGAACAGCAGATGATGAAGATGAGGCTGTCACATGATCA 1931
DB	1	AGGTAATAATCAATATGATGGAACAGCAGATGATGAGGAGGCTGTCACATGATCA 60
QY	1932	AGCCATCAGACAGGCTGCTGAAGCTGCTTAACTCTCTCAATACACATCCATCTCAT 1991
DB	61	AGCATACAGGCGAGCGCTTGAGCGTGAAGGAGGCTGCTTAAACGACCATTCCTCTT 120
QY	1992	TCAATTCGCTGGAACATTTTAATCATTTACTGCTGTCTGAATAATGATGATGATAAAGT 2051
DB	121	TCACTTCGCTGGAACATTTTAATCATTTACTGCTGTCTGAATAATGATGATGATAAAGT 180
QY	2052	AGCAGAAGCTGCACCTACAAATTTTCAAAAACACAGGAAGCAAAATGAAGGATTTTCG 2111
DB	181	GGCAGAAGCTGCACCTACAAATTTTCAAAAACACAGGAAGCAAAATGAAGGATTTTCG 240
QY	2112	ACACATCAGATCAGCCTTCTGCTTCCTGTTTATCATCAACAATCTAATAAAGACCCCGC 2171
DB	241	TCAATCAGATCAGCCTTCTGCTTCCTGTTTATCATCAACAATCTAATAAAGACCCCGC 300
QY	2172	TCAAGCCAAATATGATTCATTTGATCCATGCGATATATTTTCTAGTAAAGACCCAGTT 2231

Db	301	GCAGGCCAAATACGGCATTCATTTGATTCATGCCATTTTCTTGTATTAAGAAACCAGTT	360
Qy	2232	TGCACAGATATTGAGCCTTGCAATAGAGCCTAGATCCAAAGCAACCTGGAACATCTCAT	2291
Db	361	TGCACAGATATTGAGCCTTGCAATAGAAAGCCTAGATCCAAAGCAACCTGGAACATCTCAT	420
Qy	2292	AACACATTTGGTACTATTGGTCAATTTGCTCTCTTGCACCTGATCAATTTGCTGCTCC	2351
Db	421	AACCCCTGGTCACTATTGGTCAATTTGCTCTCTTGCACCTGATCAATTTGCTGCTCC	480
Qy	2352	TTGGAAATCTGGTAGTACTTTCATTTGTAAGATCTTTCATGAATGATGCGCTCC	2411
Db	481	TCTGAGCTTTTGGGGCACTTTCATTTGTAAGACCTCTCTCATGAATGACCGGCTTCC	540
Qy	2412	AGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAGAAAGATATCTCTGAGACAATGAT	2471
Db	541	AGGAAAAAGACAACTAAAGCTTTGGGTTCCAGATGAGAAAGTCTCACCTGAGACAATGAT	600
Qy	2472	CAAAATTCAGGCTATTAAATGATGCTGATGCTACTTGTGAATGAAAAATTAATCACAG	2531
Db	601	CAAAATTCAGGCTATTAAATGATGCTGATGCTACTTGTGAATGAAAAATTAATCACAG	660
Qy	2532	TAAATCAGAACTTTACCTTAAGATTTGCTAAACAACATATTGATAGTGAAGACTT	2591
Db	661	TAAATCAGAACTTTACCTTAAGATTTGCTAAACAACATATTGATAGTGAAGACTT	720
Qy	2592	GACAGAACAGGGAAATTTAGTAAACAGATATGTACAGTGTGAGACTTGTCTGGAG	2651
Db	721	GACAGAACAGGGAAATTTAGTAAACAGATATGTACAGTGTGAGACTTGTCTGGAG	780
Qy	2652	TGCAATGTGAAGCTGGACAAAGAACCTGTTTACATAAATCATCAATTAAGACATA	2711
Db	781	TGCAATGTGAAGCTGGACAAAGAACCTGTTTACATAAATCATCAATTAAGACATA	840
Qy	2712	TCACATGATGCTATGATCAACAGATGAATGCTAAGTAAGTAAGTAAGTGTGCCCA	2771
Db	841	CCAGCTGTGCTATTAGCATCAATGATGATGATTAACAAGTCAAGGAGTGTTCGCCCA	900
Qy	2772	GAAACTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAATATATGCAATCTGTGC	2831
Db	901	GAAACTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAATATATGCAATCTGTGC	960
Qy	2832	CCTTTGTCACAAAGATCCTGTAAAGAGAGAAAGCCTATGCTAGACATGTTTGGTGA	2891
Db	961	TCTTTGTGCCAAAGACCTGTGGAAGAGAGAGGCGACCATGCTAGACAGTGTGTGTA	1020
Qy	2892	AAATTAATTAATTAAGGCGAGATCTGAAGCAGCATGACAGTGTAGTGAATAATTTT	2951
Db	1021	GAACATCACTGTGAGAGGAGAGTACCTGAACAGCATGACAGTGTAGTGAATAATTTT	1080
Qy	2952	GTCCTTCTACAGATGATGTTTCCATATACATTTACCTTTTGGACATGACCCGAG	3011
Db	1081	GTCCTTCTTACCAAGATATGTTGTTCCATATACATTTACCTTTTGGACATGACCCGAG	1140
Qy	3012	TTATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGCTTTGGTGT	3071
Db	1141	TTATGTCAAAGTACAGATATTGAACAACCTTAAGATGTTAAAGATGCTTTGGTGT	1200
Qy	3072	TCTGAAATATTAAATGCTTAAATTAAGAAATTAACAGTCAAGCTTTTATCAGAAAGAT	3131
Db	1201	TCTGAGATATTGATGGCTTAAATTAAGAAATTAACAGTCAAGCTTTTATCAGAAATG	1260
Qy	3132	AGAAAAATTTAAACAAACAAAGATGCCAAAGGACAGATGATGAAAAATGATGAAAA	3191
Db	1261	AGAAAAATTTAAACAAACAAAGATGCCAAAGGACAGATGATGAAAAATGATGAAAA	1320
Qy	3192	ACTGTACAGCTGTGTGATGTTGCCATGAATATCATCTGCAAGAGATATACACTAG	3251
Db	1321	ATTGTACAGCTGTGTGATGTTGCCATGAATATCATCTGCAAGAGATATACACTAG	1380
Qy	3252	TTTGAATCTCTTAAAGACCCGCTACTACAGCTGTCTTCTCACTCAACCTGCACAGAA	3311
Db	1381	CCTGAGCTCTCTTAAAGACCCCGCTCTCACAGCTGGTTTTTCACCCAGCCTGCACAGAA	1440
Qy	3312	TTTCAGTAACACCAAAAAATTAATCTGCTCTCTGAAATGAATATCTTTTCACTCTGAAA	3371
Db	1441	TTTCAGTAACACCAAAAAATTAATCTGCTCTCTGAAATGAATATCTTTTCACTCTGAAA	1500
Qy	3372	ACCTAAACCAACCAATGTTCTTAGAGAGCTGTTAAACAGGCACCTTTCATACAGAGCA	3431
Db	1501	ACCTAAACCAACCAATGTTCTTAGAGAGCTGTTAAACAGGCACCTTTCATACAGAGCA	1560
Qy	3432	ATCTAGACCAAAATCATACAGATGAAACTGTAAAGCAATGCAACAGCAGCTCAATTC	3491
Db	1561	GTCTAGACCAAAATCATACAGATGAAACTGTAAAGCAATGCAACAGCAGCTCAATTC	1620
Qy	3492	AAGCTCTCTGGAAGATTAAGGGAGGCTTGATATGTTCTGAATGATCAGCTGAAA	3551
Db	1621	AAGCTCTCTGGAAGATTAAGGGAGGCTTGATATGTTCTGAATGATCAGCTGAAA	1680
Qy	3552	TGAAGATTACCAATGCTTTCACCTTGGCCGGGAAAAAAGTGCAGAGAGAGAGCTC	3611
Db	1681	TGAAGATTACCAATGCTTTCACCTTGGCCGGGAAAAAAGTGCAGAGAGAGAGAGCTC	1740
Qy	3612	TGATCTTTAAGGTTGTAATTTGAGAACCTTAGAGCAGAGAAAAAACGCCCTGCACAG	3671
Db	1741	TGATCTTTAAGGTTGTAATTTGAGAACCTTAGAGCAGAGAAAAAACGCCCTGCACAG	1800
Qy	3672	ACAGGAGAGAAATTAAGTATGATGATGACTTGTGACTAAGTTGGTACAGGAACGAACCTTA	3731
Db	1801	CCCTGMAAGAAATTAAGTATGATGATGATGACTAAGTTGGTACAGGAACGAACCTTA	1860
Qy	3732	AGGCACTAGGCAAGTCTGGAAGAAAGGCGCATAGGCTTCAAGATCTGAATGCAACACATG	3791
Db	1861	AGGCACTAGGCAAGTCTGGAAGAAAGGCGCGTACAGCTTCAAGATCTGAAGTCAAGAG	1920
Qy	3792	GCTTGAGGAAAAAGGCTCAAAAGAAATATATTAGAAATGAAGATGAACGAATATGCTC	3851
Db	1921	GCTTGAGGAAAAAGGCTCAAAAGAAATATATTAGAAATGAAGATGAACGAATATGCTC	1980
Qy	3852	GCCAAAAAAGGTTAAAGAGGCCGACACCAAAACCTTGTGTGAGATACCAAAAGA	3911
Db	1981	ACCAAAAAAGGTTAAAGAGGCCGACACCAAAACCTTGTGTGAGATACCAAAAGA	2040
Qy	3912	AGAGCAACAAATGAAGAACTTTAAAGAAAGAAAGCAAAAAAATCTGACCTCCAGCAC	3971
Db	2041	AGAGCCCAATGAAGAAATATCCAAAGAAAGAAAGCAAAAAAATCTTGTACCTCTGTGT	2100
Qy	3972	AGAGGAGAGAGAGAGAGAGAGCAAAATGCGAATACGGAACGAAGTCCAAAGCA	4031
Db	2101	AGAGGAGAGAGAGAGAGAGAGCAAAATTTGAAACACAGAACATTAAGTCAAAAAGCA	2160
Qy	4032	ACAGCAGCAGTGTCAAGAGAGCAGCAGCAGCAGCAAGTCTGTAATCTAGTGCAT	4091
Db	2161	ACAGCAGCAGCAGTGTCAAGAGAGCAGCAGCAGCAGCAAGTCTGTAATCTAGTGCAT	2220
Qy	4092	TGAATCCACAGCTCCACACACAGAGAAAGAGAGAGAGACCATCAAAAAGCCTACAC	4151
Db	2221	TGAATCCACAGCTCCACACACAGAGAAAGAGAGAGAGACCATCAAAAAGCCTACAC	2280
Qy	4152	ATCACACCAAAAAAAA 4169	
Db	2281	ATCACACCAAAAAAAA 2298	
RESULT 2			
AK041682 2899 bp mRNA linear HTC 05-DEC-2002			
LOCUS AK041682			
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched			
library, clone:630029M15 product:49110.1.2 (ANDROGEN-INDUCED			
PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)			
(FRAGMENT) homolog [Homo sapiens], full insert sequence.			
ACCESSION AK041682			
VERSION AK041682.1 GI:26334660			
KEYWORDS HTC; CAP trapper.			

[illegible]

OY 1730 AAGTGTGTGCGTGAATACTAGAGATTGGGCAACCCCAACAGCCTACAAATCCCT 1789
 DB 259 AAGCGTGTGCGTGAATACTAGAGATTGGGCAACCCCTACAGAGCCTACAAACCCCT 318
 OY 1790 TCTGGAATATGATCAAGTTTCTCTTGAGAGATGACACTGTGCACATAGATACGAAT 1849
 DB 319 TCTTGAATATGATCAAGTTTCTCTTGAGAGATGATGCTGTACACATAGACATGAT 378
 OY 1850 CTATGAGTGTCTTATTTAAACAAGTGAACAAATCAATAGATGAACAGACATGATGAAG 1909
 DB 379 CCATCAGTGTCTTATTTAAACAAGTGAACAAATCAATAGATGAACAGACATGATGAAG 438
 OY 1910 ATGAGAGTGTTCACAACTGATCAAGCCATCAGAGAGGCTTGAAGTGTGATGATGAT 1969
 DB 439 ATGAGAGTGTTCACAACTGATCAAGCCATCAGAGAGGCTTGAAGTGTGATGATGAT 498
 OY 1970 CATTTACACATCCCATCTCATTTCTGCTGGAACAATTTGAATCATTTACTGCTGTGC 2029
 DB 499 CTTTACGACCCCTCTCTCTCTCTGCTGGAAGCTTTGAGTCCCTCTGCTGTGCT 558
 OY 2030 TGAATAATGATGATGAAAAATAGAGAGAGCTGCTCAAAATTTTCAAAAAACAGAGAA 2089
 DB 559 TGAAGATGATGATGAGAGAGAGCTGCTCAAAATTTTCAAAAAACAGAGAA 618
 OY 2090 GCAAAATTTGAAGAGATTTTCCACACATCAGATCAGCTTGCCTGCTTTTACATCA 2149
 DB 619 GCAAAATTTGAAGAGATTTTCCCTCCTCCTCCTCCTGCTTCTCTCTCTCTCTCTCA 678
 OY 2150 AATCTAAAAAGAT 2209
 DB 679 AATCTAAAAAGAT 738
 OY 2210 TTTCTAGTAAAGAT 2269
 DB 739 TTTCTAGTAAAGAT 798
 OY 2270 CAAGAACCTGGAACATCTCATTAACACCATTTGCTACTATGCTATATGCTCTCTG 2329
 DB 799 CAAGAACCTGGAACATCTCATTAACACCATTTGCTACTATGCTATATGCTCTCTG 858
 OY 2330 CACCTGATCAATTTGCT 2389
 DB 859 CACCTGATCAATTTGCT 918
 OY 2390 TTTCTAGTAAAGAT 2449
 DB 919 TTTCTAGTAAAGAT 978
 OY 2450 AAGTATCTCTGAGACATGCTCAAAATTTAGAGCTATTAATGATGCTGCTGCTGCT 2509
 DB 979 AAGTATCTCTGAGACATGCTCAAAATTTAGAGCTATTAATGATGCTGCTGCTGCT 1038
 OY 2510 TTGGAATGAAAAATATATACAGTAATACAGAACTTACCTTAAGATTGCTTAACAGAA 2569
 DB 1039 TTGGAATGAAAAATATATACAGTAATACAGAACTTACCTTAAGATTGCTTAACAGAA 1098
 OY 2570 TATTCATATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2629
 DB 1099 TATTCATATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
 OY 2630 GTCTGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2689
 DB 1159 GCTGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218
 OY 2690 AATATCATCATATGAGAACATATACAGTATGCTATGATCAATCAAGATGATGCTATC 2749
 DB 1219 AATATCATCATATGAGAACATATACAGTATGCTATGATCAATCAAGATGATGCTATC 1278
 OY 2750 AAGTAAAGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2809
 DB 1279 AAGTAAAGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338

OY 2810 TTGAGTATATGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2869
 DB 1339 TTGAGTATATGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1398
 OY 2870 ATGCTAGGCAATGTTTGGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 2929
 DB 1399 ATGCTAGGCAATGTTTGGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1458
 OY 2930 CAGCTGTAGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2989
 DB 1459 CAGCTGTAGTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1518
 OY 2990 ACCTTTGGACATGACCCAGATATGCTCAAAAGTACAGAGATATGGAACATTAAGATG 3049
 DB 1519 ACCTTTGGACATGACCCAGATATGCTCAAAAGTACAGAGATATGGAACATTAAGATG 1578
 OY 3050 TTAAGAAATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3109
 DB 1579 TGAAGAAATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1638
 OY 3110 ACGCTTTATCAGAAAGATGCTAGAAATATTAATTAATTAATTAATTAATTAATTAAT 3169
 DB 1639 ACGCTTTATCAGAAAGATGCTAGAAATATTAATTAATTAATTAATTAATTAATTAAT 1698
 OY 3170 ATGATGCAAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 3229
 DB 1699 ATGATGCAAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 1758
 OY 3230 TGTCAAGAGTATCATATGATGTTGAAATCTCTAAAGACCCGCTACTACAGCTGCT 3289
 DB 1759 TGTCAAGAGTATCATATGATGTTGAAATCTCTAAAGACCCGCTACTACAGCTGCT 1818
 OY 3290 TCTTCACTCAACCTGACAAAGATTTCAATACACCAAAATTAATTAATTAATTAATTAAT 3349
 DB 1819 TCTTCACTCAACCTGACAAAGATTTCAATACACCAAAATTAATTAATTAATTAATTAAT 1878
 OY 3350 AATCATTTTTCACCTGCTGAAACCTTAACCAACCAATGTTCTAGAGCTGTTAAACAGC 3409
 DB 1879 AATCATTTTTCACCTGCTGAAACCTTAACCAACCAATGTTCTAGAGCTGTTAAACAGC 1938
 OY 3410 CACTTTTATCAGAGAGCAAGCAATCTCAGACCAATTAATTAATTAATTAATTAATTAAT 3469
 DB 1939 CACTTTTATCAGAGAGCAAGCAATCTCAGACCAATTAATTAATTAATTAATTAATTAAT 1998
 OY 3470 ATGCAAG 3529
 DB 1999 ATGCAAG 2058
 OY 3530 CTGAATATGATCAG 3589
 DB 2059 CTGAATATGATCAG 2118
 OY 3590 AAAGTGACAG 3624
 DB 2119 AAAGTGACAG 2153
 RESULT 3
 LOCUS AK086753
 DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:0930049H23, product:49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens), full insert sequence.
 ACCESSION AK086753
 VERSION AK086753.1 GI:26103739
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Carninci, P. and Hayashizaki, Y.

JOURNAL MEDLINE PUBMED	99279253 10349636	High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
REFERENCE	2	
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL MEDLINE PUBMED	20493574 11042159	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	3	
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, T., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, Y., Tanaka, T., Matsuzawa, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer	
JOURNAL MEDLINE PUBMED	20530913 11076861	Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	4	
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kodato, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schiraldi, L. M., Stubbli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamaya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Rongwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S., and Hayashizaki, Y.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
JOURNAL MEDLINE PUBMED	21085660 11217851	Nature 409 (6821), 685-690 (2001)
REFERENCE	5	
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL MEDLINE PUBMED	Nature 420, 563-573 (2002)	
REFERENCE	6 (bases 1 to 4273)	
AUTHORS	Adachi, J., Aizawa, K., Akiyama, T., Hara, A., Hashizume, M., Fukuda, S., Furuno, M., Hanagata, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocanti, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-APR-2002)	
TITLE	Submitted (16-APR-2002) yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,	

FEATURES	source
COMMENT	<p>Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9216, Fax: 81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site for further details.</p> <p>URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers</p>
FEATURES	<p>1. .4273</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="FANTOM_DB:D930049H23"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="D930049H23"</p> <p>/tissue_type="head"</p> <p>/clone_id="RIKEN full-length enriched mouse cDNA library"</p> <p>/dev_stage="15 days embryo"</p> <p><1. 2245</p>
misc_feature	<p>/note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog (Homo sapiens) (SPTR1Q96K4, evidence: FASTV, 92.3%ID, 100%length, match=1430)</p> <p>putative"</p>
BASE COUNT	1363 a 840 c 899 g 1171 t
ORIGIN	
Query Match	39.7%, Score 1655.2, DB 11, Length 4273;
Best Local Similarity	88.8%, Pred. No. 9.2e-274;
Matches 1836; Conservative	0; Mismatches 228; Indels 4; Gaps 4;
OY	2104 GATTTTCACACATCAGATGACGCTTCGCTTCCTTTTACATCAACAATCT-AAAAAGG 2162
DB	1 GATTTCCTTCACATCAATCAAGCTTCCTTCCTTACTACTACATCAACAATAAAGG 60
OY	2163 ACCCCCCCGTCAGCCCAATATGCAATTCATTTATCATGCGATATTTCTACTAAGA 2222
DB	61 ACCACCTTCGCGAGCCCAATACCCCATTCATTTATCATGCCATATTTCTACTAAGA 120
OY	2223 GACCACGTTTCCACAGATATTTGAGCGCTGSCATAAAGAGCCTAGATCCAAAGCCGGA 2282
DB	121 AACCCAGTTTCCACAGATATTTGAGCCCTGSCATAAAGAGCCTAGATCCAAAGCCGGA 180
OY	2283 ACATTCATCAACACATTTGTTACTATTTGTCATATTTGCTCTCCCTTCACACTGATCAAT 2342
DB	181 GCATCTTATTAACCCCTTCGTCACATTTGTCATATTTGCTCTTCGCACACTGATCAGTT 240
OY	2343 TGTGTCGTCCTGGAAATCTGGGATGAGTACTCTTATTTGTAAGAAATCTTCCATGAATGA 2402
DB	241 CGGTGTCCTCTGGAAGTCCTTTGGGGCACTTTTATTTGTGAAGAACCTCCATCAATGAATGA 300
OY	2403 TCGGCTTCCAGGAAAAAGACACTTAACTTTGGGTTCCAGATGAAGAAGTATCTCCCTGA 2462
DB	301 CCGGCTTCCAGGAAAAAGACACTTAACTTTGGGTTCCAGATGAAGAAGTATCTCCCTGA 360
OY	2463 GACATATGTCAAAATTCAGGCTATTTAAATGATGTTGATGCTTACTTGGAAATGAATAA 2522
DB	361 GACATATGTCAAAATTCAGGCTATTTAAATGATGTTGATGCTTACTTGGAAATGAATAA 420
OY	2523 TAATTCACAGTAAATCAGGAACCTTCACTTAAAGTGTGTATCAACATATTTGCTGATAGGA 2582
DB	421 TAATTCACAGTAAATCAGGAACCTTCACTTAAAGTGTGTATCAACATATTTGCTGATAGGA 480
OY	2583 TGGAGACTTGCAGAACAGGGGAAATTTAGTAAACAGATATGTCACGTCAGACTTGC 2642
DB	481 TGGGAGATTTGCAGAACAGGGGAAATTTAGTAAACAGATATGTCACGTCAGACTTGC 540
OY	2643 TGTCTGGAGTCTCTTTGTGAAGCTGGCAACAAGAACCTGTTACATGAATCATACATTT 2702

Db 541 TCTGGGAGTCTATTGTGAAGCTGGCAGAGGCCCTGTTACCAAGATCATTTACACT 600
 Oy 2703 AGAACAAATATPAGCATGTATGTAGCATATCAAGCATGAATGCTCTCAAGTAAGCAAGT 2762
 Db 601 GGAGCAGTACAGAGCTGTGTGATGATCCATCAATGATGATGTATCAAGTACAGCAGGT 660
 Oy 2763 GTTTGCCGAGAACTTCACAAAGGCCCTTCCGTTTACGGCTTCCACTT - GAGTATATGG 2821
 Db 661 GTTGGCTCAGAACTTCACAAAGGCCCTTCCGCTTACGGCTTCCCTTGGAGTACATGG 720
 Oy 2822 CAATCTGTGCCCTTTGTGCAAAAGATCTGTAAAGAGAGAGAAGAGCATCTAGTCAAT 2881
 Db 721 CCATCTGTGCTCTTGTGTGCCAAAGACCTGTGAAAGAGAGGAGGAGCCATGCTACAGT 780
 Oy 2882 GTTGTGTGAATAAT 2941
 Db 781 GCTTGTGTGAATAAT - CCCGTGAGAGGAGGATACCTTAACAGCATGCTACCTGTATGTG 839
 Oy 2942 AAAAATTTATTTCTCTTCTACAGAGTATGTTGTTCCATATACATTTACCTTTTGGCAC 3001
 Db 840 AAAAATTTATTTCTCTTCTTACAGAGTATGTTGTTCCATATACATTTTGGCAC 899
 Oy 3002 ATGACCCAGATTTATGTCAAGATACAGATATTTGAACAATTAAAGATTTAAAGATGTC 3061
 Db 900 ATGACCCAGATTTATGTCAAGATACAGATATTTGAACAATTAAAGATTTAAAGATGCC 959
 Oy 3062 TTTGGTTTGTCTGGAATATTTAATGCTAAATAATGAAATAATACAGTACGCTTTTATCA 3121
 Db 960 TTTGGTTTGTCTGGAATATTTAATGCTAAATAATGAAATAATACAGTACGCTTTTATCA 1019
 Oy 3122 GAAAGATGTAGAAAATTTATTAACAAACAAAGATGCCAAGAGACAGATGTGCAAAA 3181
 Db 1020 GAAAGATGTAGAAAATTTATTAACAAACAAAGATGCCAAGAGACAGATGTGCAAAA 1079
 Oy 3182 TGAATGAAAACCTGACACGTGTGTGATGTTGCCATGATCATCATGTCACAAAGATA 3241
 Db 1080 TGAATGAAAACCTGACACGTGTGTGATGTTGCCATGATCATCATGTCACAAAGATA 1139
 Oy 3242 CTACATACAGTTTGGAAATCTCTTAAGAACCCGGTACTACAGCTGTTTCTTCACTAC 3301
 Db 1140 CCACGTACAGCTGTGAGTCTCTTAAGAACCCGGTCTGCTGCAAGCTGTTTCTCACTAC 1199
 Oy 3302 CTGACAGAAATTTAGTACACACAAATTTATCTGCCCTCGAAGATGAATCATTTTCA 3361
 Db 1200 CTGACAGAAATTTAGTACACACAAATTTATCTGCCCTCGAAGATGAATCATTTTCA 1259
 Oy 3362 CTCTCGAAAACCTTAACCAACCAATGTTTGAAGAGCTGTTTAACAGGCACTTTCATCAG 3421
 Db 1260 CTCTCGAAAACCTTAACCAACCAATGTTTGAAGAGCTGTTTAACAGGCACTTTCATCAG 1319
 Oy 3422 CAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAACGTGAAGCAATGCAAGCAGCA 3481
 Db 1320 CAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAACGTGAAGCAATGCAAGCAGCA 1379
 Oy 3482 GCTCAAAATCCAAAGCTCTCTGGAAGATTAAGGGAGGCTGTAGTCTGAAATGATGTC 3541
 Db 1380 GCTCAAAATCCAAAGCTCTCTGGAAGATTAAGGGAGGCTGTAGTCTGAAATGATGTC 1438
 Oy 3542 ACAGTGAATTAAGATTTACCAATGCTCTTACCTTTGCGGGGAAAAAAGTACAGCA 3601
 Db 1439 ACAGTGAATTAAGATTTACCAATGCTCTTACCTTTGCGGGGAAAAAAGTACAGCA 1498
 Oy 3602 GAGGAGCTCTGATCTGTAGAGTCTGAATGGAAGCCCTAGAGCCGAGGAAAAAAGCC 3661
 Db 1499 GAGGAGCTCTGATCTGTAGAGTCTGAATGGAAGCCCTAGAGCCGAGGAAAAAAGCC 1558
 Oy 3662 CCGTCAAGAAACAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATG 3721
 Db 1559 CCGTCAAGAAACAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATG 1618
 Oy 3722 AGAAACCTAAAGGAGTACGGAAGTGGAAAAAGGCAATACGGCTTCAGATCTGATG 3781
 Db 1619 AGAAACCTAAAGGAGTACGGAAGTGGAAAAAGGCAATACGGCTTCAGATCTGATG 1678

Oy 3782 AACAGACTGGCTGAGGAAAAAGAGCTCAAGAGATATATATGAAATGAGATGAC 3841
 Db 1679 AGCAGACTGGCTGAGGAAAAAGAGCTCAAGAGATATATATGAAATGAGATGAC 1738
 Oy 3842 AGAATAGTCCGCAAAAAAGGTTAAAGAGGCCGACCAACCAAACTCTTGGTGGAGTA 3901
 Db 1739 AGAAGAGCCCAAAAAAGGCAAAAGAGGAGCCACCAAACTCTTGGTGGAGTA 1798
 Oy 3902 CACCAAAAGAGAGCCCAATTAATGAAATCTTAAAAAAGAGCAAAAAAATCTGCAC 3961
 Db 1799 CATGCAAGAGAGAGCCCAATTAATGAAATCTTAAAAAAGAGCAAAAAAATCTGTAC 1858
 Oy 3962 CTCACAGCAG 4021
 Db 1859 CTCCTGTATGAGACAGCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
 Oy 4022 CCAAAAGCAACAGCAGAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4081
 Db 1919 CAAAGAGCAACAGCAGAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
 Oy 4082 CTAGTCAATTTGAATCCACACAGTCCACACAGAGAGAGAGAGAGAGAGAGAGAG 4141
 Db 1979 CAAGTGAGTGAATCCACACAGTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
 Oy 4142 CGCATCCACATCCACCAACCAAAAAA 4169
 Db 2039 CACCATCCACATCCACCAACCAAAAAA 2066

 RESULT 4
 AK032384
 LOCUS
 DEFINITION
 Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430530P15 product:49310.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert sequence.
 ACCESSION
 AK032384
 VERSION
 AK032384.1 GI:26082814
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
 REFERENCE
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 MEDLINE
 PUBMED
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20493574
 MEDLINE
 PUBMED
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishikawa, K., Katsunuma, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 PUBMED
 11076861
 4
 REFERENCE

OY	3173	TTGCAAAATTAAGAAAAACTGTACACTGTGTGTGATGGTCCATGAATATCAATGT	3323
Db	361	ATTACAAAAATGATGAAAAATTGTACACCGTGTGTGATGGTCCATGACATCATCTGT	420
OY	3333	CAAAGAGTACTACATACAGTTTGGAAATCTCCTTAAAGACCCGGTACTACAGCTGTTCT	3392
Db	421	CCAAGAGCACACGTACAGCCTGGAGTCTCTTAAGAGACCCGCTGTCAGACTCGTTTT	480
OY	3293	TCACGTCAACCTGCAGAAAGATTCTGTAAACCCAAAAATTATCTGCGCTCTGAATGAAT	3352
Db	481	TCACCCAGCCCTGACAGAGATTTTGTAAACACCAAAAAATTACTCTCTCCAGAAAAAT	540
OY	3353	CATTTTTCACCTCCGAGAAAACCTTAAACCAACATATGTTCTAGAGAGCTGTTAAACAGAC	3412
Db	541	CATTTTTCACCTCCGAGAAAACCTTAAACAGCAATGTTCTTCGAGCAGTTAATTAAGCCAC	600
OY	3413	TTTCATCAGCAGCCAGCAACATCTCAGACCAATATCATCAGANTGAAACGTGAACATG	3472
Db	601	TTTCATCAGCAGCCAGCAACAGTCTCAGACCAATATCATCAAGANTGAAACTGTGACAAAG	660
OY	3473	CAACGAGAGGTCAAAATCCAAAGCTCCTCTGGAAGAAATTAAGGGAGGCTGTGATGTTCTG	3532
Db	661	CAACGAGAGGTCTCAACCCAAAGCTCCTCTGGAAGATCAAGGGAGGCTGTGATGTTCTG	720
OY	3533	AAATGGATCAGCAGTGAATAATGAATATTACAAATGTCTTCACTTGTGCGGGGAAAAAA	3592
Db	721	AAATGGATCAGCAGTGAATAATGAATATTACAAATGTCTTCACTTGTGCGAGAAAAAAA	780
OY	3593	GTGACAAAGAGAGCAGCTGTGATCTTGTAAAGTCTGAATTTGGAGAAAGCCTTAGAGGACGA	3652
Db	781	GTGACAAAGAGAGAGACCTGTATCTT-----TCGTAGTTTGGAAAGCCTTAGAGTCCGA	834
OY	3653	AAAAAAGCCCGTCAAGAAAGAGGAAATTAGATATGATGATGACTTGCTAATGTTG	3712
Db	835	AAAAAAGCCCTGTCAAGACACCTTGAAGGAAATTAGATATGATGATGACTTACTAATGTTGG	894
OY	3713	TACAGGAACAGAAACCTTAAAGCAGTCAAGGAGGCGGAAAAAGAGCCGTACACTTCAG	954
Db	895	TACAGGAACAGAAACCTTAAAGCAGTCAAGGAGGCGGAAAAAGAGCCGTACACTTCAG	
OY	3773	AATCTGATGAACAGCAGTGTGGCTGTGAGAAAAAGAGCTCTCAAGAAATATATTACAAATG	3832
Db	955	ACTGAGAGACAGCAGCAGTGTGGCTGTGAGGAAAGAGGACCAAAAGAGAGTCTCTGAAATG	1014
OY	3833	AAGATGAACAGAAATAGTCCGCGCAAAAAAGGTTAAAGAGGCGCGACACCAAAACCTCTTG	3892
Db	1015	AGATGTAGCAGAAACAGCCACCAAAAAAGGCCAAAGAGCGACCAAAACCTCTTG	1074
OY	3893	GTGAGAGTACACCAAAAGAGAGCCCAACATGAAAACCTTCTTAAAAAGGAACAAAAAA	3952
Db	1075	GTGAGAGGAGACATCGAAGAGAGAGCCCAACATGAAAACATCCAAAGAAAGAAACAAAGAAA	1134
OY	3953	AATCTGAGCCTCCAGCACCCAGAGAGAGGAGAAAGAAAGACAAAGTGGAAATACGG	4012
Db	1135	AACTTGTATCTCTGTAGTATGACGACGATGAAGAAAGAAAGACAAATTTGAAAACACAG	1194
OY	4013	AACAGAAATCTCAAAAAGCAAAACAGACAGATGTCTCAAGAGAGAGCAGCAGAGACAGAT	4072
Db	1195	AACATTAATCTCAAAAAGCAAAACAGACAGCCCAATCAAGAGAGGACACACAGAGACAGAT	1254
OY	4073	CTCCTGAATCTAGTGCAAATTGAAATCCACACAGTCCACACCAAGAAAAGAGCAGAGAAAGAC	4132
Db	1255	CTCCTGAAGAACAGAGTGAATTAATTCACACAGTCCACACCAAGAAAAGAGAGAGAAAGAC	1314
OY	4133	CATCAAAAAGCCCATCCACATCACACCAACAAAAAAA	4169
Db	1315	CATCAAAAAGCACATCACATCACACCAACCCCAAAAA	1351

RESULT 6					
BF509252					
LOCUS	BF509252	791 bp	mRNA	linear	EST 06-DEC-2000
DEFINITION	UT-H-B14-aow-c-07-0-UT.s1	NCI_CGAP_Sub8	Homo sapiens	cdna	clone

Accession	Version	Keywords	Source	Organism	Reference Authors Title	Journal Comment
IMAGE:3086220	3'	RNA sequence.				
Bf509252						
Bf509252.1	GI:11592550					
EST.						
Homo sapiens (human)						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.						
NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.					
1 (bases 1 to 791)						
NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.					
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
Tumor Gene Index						
Unpublished						
Contact: Robert Strausberg, Ph.D.						
Email: cga@b5-nci.nih.gov						
The sequence contained an oligo-dT track that was present in the						
oligonucleotide that was used to prime the synthesis of first						
strand and cDNA and therefore this may represent a bonafide poly A						
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution::						
NCI-CGAP clone distribution information can be found through the						
I.M.A.G.E. Consortium/LLNL at:						
www-bio.llnl.gov/bbrr/image/image.html						
Seq primer: M13 Forward						
PolyA=Yes.						
Location/Qualifiers						
1..791						
/organism="Homo sapiens"						
/mol_type="mRNA"						
/db_xref="taxon:9606"						
/clone IMAGE:3086220"						
/lab_host="DH10B (Life Technologies)"						
/clone_id="NCI-CGAP-Sub8"						
/note="Vector: pRT3D-Pac (Pharmacia) with a modified						
polylinker; Site_1: Not I; Site_2: Eco RI; NCI-CGAP-Sub8						
is a subcloned library derived from NCI-CGAP Sub5. The						
NCI-CGAP-Sub8 library had 2.5 million recombinants. A						
single-stranded DNA preparation of NCI-CGAP-Sub5 was used						
as a tracer in a subtractive hybridization with a driver						
comprising: a pool of clones from NCI-CGAP-Sub5 (IMAGE						
clone ids 2732833-2737415, 3068040-3069191; 25% of the						
driver population), a pool of clones from NCI-CGAP-Sub4						
(IMAGE clone ids 2723592-2729326; 25% of the driver						
population), NCI-CGAP-Sub6 (pool AIF-AJF, IMAGE ids						
2728969-2733190; 25% of the driver population) and						
NCI-CGAP-Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550						
; 25% of the driver population). Subtraction was						
performed as previously described [Bonaldi, Lennon &						
Soares (1996): Normalization and Subtraction: Two						
Approaches To Facilitate Gene Discovery. Genome Research						
6 791-806.						
TAG_LIB=NCI-CGAP-Lu5						
TAG_TISSUE=Lung						
TAG_SEQ=CAAC"						
BASE COUNT	270 a	151 c	144 g	226 t		
ORIGIN						
Query Match	17.2%	Score 716;	DB 10;	Length 791;		
Best Local Similarity	99.6%	Pred. No. 1e-112;				
Matches 749;	Conservative 0;	Mismatches 0;	Indels 3;	Gaps 3;		
2680	TGTTACCTGTAATCATCATCATTTAGAACAAATATGATGTGATTCATCAACAT	27399				
13	TGTTACCTGTAATCATCATCATTTAGAACAAATATGATGTGATTCATCAACAT	72				
2740	GAATGCTATCAAGTAGAACAAGTGTGGCCAGAAACTTCACAAAGGCTTTCCGTTTA	27999				
73	GAATGCTATCAAGTAGAACAAGTGTGGCCAGAAACTTCACAAAGGCTTTCCGTTTA	132				
2800	CGGCTCCACCTGAGTATATGAGCAATCGTCCCTTTGTGCAAAAAGATCCTGTAAGAG	2859				
133	CGGCTCCACCTGAGTATATGAGCAATCCTGCTCCCTTTGTGCAAAAAGATCCTGTAAGAG	192				
2860	AGAAAGAGCTCATGCTAGGCAATGTTTGGTGAATAATATGTAAGGCGGAGATATCTG	2919				

```

|||||
193 AGAAGAGCTCATGTAGGCAATGTTGGTGAATAATTAATGTAAGGGGGGATATCTG 252
QY 2920 AAGCAGCATGCAGCTGTTAGTGAATAATTTATGCTCTTACACGAGTATGTTGCCA 2979
Db |||||||
253 AAGCGCATGCGAGCTGTAGTGAATAATTTATGCTCTTACACGAGTATGTTGCCA 312
QY 2980 TATTAATTTACCTTTGGCAGATGACCCAGATTATGCAAGTACAGATATTGACAA 3039
Db |||||||
313 TATTAATTTACCTTTGGCAGATGACCCAGATTATGCAAGTACAGATATTGACAA 372
QY 3040 CTTAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATTAATGCTAAATGAA 3099
Db |||||||
373 CTTAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATTAATGCTAAATGAA 432
QY 3100 AATAACAGTACAGCTTTTATTCGAAAAGATGTAAGAAATTTAAACAACAAGATGCC 3159
Db |||||||
433 AATAACAGTACAGCTTTTATTCGAAAAGATGTAAGAAATTTAAACAACAAGATGCC 492
QY 3160 CAAGACCATGATGATGCAAAAATGATGTAAGAAATGTAAGTACAGTGTGATGTTGCCATG 3219
Db |||||||
493 CAAGACCATGATGATGCAAAAATGATGTAAGAAATGTAAGTACAGTGTGATGTTGCCATG 552
QY 3220 AATATCATCATGCTCAAGAGTACTACATACAGTTGGAAATGCTTAAGACCCGGTACTA 3279
Db |||||||
553 AATATCATCATGCTCAAGAGTACTACATACAGTTGGAAATGCTTAAGACCCGGTACTA 612
QY 3280 CCAGTGTGTTCTTACTCAACCTGACCAAGATTTTCAGTAACACCAAAATTTATCTGCT 3339
Db |||||||
613 CCAGTGTGTTCTTACTCAACCTGACCAAG-ATTTCAGTACACCAAAATTTATCTGCT 671
QY 3340 CCTGAATGAATCATCTTTTCTCTCTCTGGAAGAACTTAACAACCAATGTTAGAGCT 3399
Db |||||||
672 CCTGAATGAATCATCTTTTCTCTCTCTGGAAGAACTTAACAACCAATGTTAGAGCT 730
QY 3400 GTTAACAAGCCACTTTCATCAGCAGCAAGCA 3431
Db |||||||
731 GTTAACAAG-CACTTTCATCAGCAGCAAGCA 761

RESULT 7
CA980171 926 bp mRNA linear EST 06-JAN-2003
LOCUS AGENCOURT_11295832 NIH_MGC_164 Mus musculus cDNA clone
DEFINITION IMAGE:30145391 5', mRNA sequence.
ACCESSION CA980171
VERSION CA980171.1 GI:27512825
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerogomphi; Muridae; Murinae; Mus.
1 (bases 1 to 926)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@strausberg.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
http://image.llnl.gov
Plate: NDAM0059 row: f column: 24
High quality sequence stop: 620.
Location/Qualifiers
1..926
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30145391"

```

```

/lab_host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_164"
/note="Vector: pCMV-Sport6.1, ccdb, Site_1: EcoRV, Site_2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp. Average insert size 1.8k bp. Priming
sequence: 5'-GACTAGTCTTCTAGTCCGACGCGCCGCTT-3'. Tissue
contributed by, David Rowe. Library constructed by Resgen,
Invitrogen Corp."

BASE COUNT      332 a      199 c      222 g      172 t      1 others
ORIGIN

Query Match      16 68; Score 691.6; DB 14; Length 926;
Best Local Similarity 87.68; Pred. No. 1.5e-108;
Matches 790; Conservative 0; Mismatches 109; Indels 3; Gaps 3;

QY 3042 TAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATTAATGCTAAATGAAA 3101
Db |||||||
25 TAAAGATGTTAAAGATGCTTTGTTGTTGTTGCAAAATTAATGCTAAATGAAA 84
QY 3102 TAACAGTACAGCTTTTATTCGAAAAGATGTAAGAAATTTAAACAACAAGATGCCA 3161
Db |||||||
85 CAACAGCATGATTTATTCGAAAAGATGTAAGAAATTTAAACAACAAGATGCTCA 144
QY 3162 AGGACCATGATGTAAGAAATGTAAGAAATGTAAGTACAGTGTGATGTTGCCATGAA 3221
Db |||||||
145 AGGACCATGATGTAAGAAATGTAAGAAATGTAAGTACAGTGTGATGTTGCCATGAA 204
QY 3222 TATCATCATGCTCAAGAGTACTACATACAGTTGGAAATGCTTAAGACCCGGTACTA 3281
Db |||||||
205 CATCATCATGCTCAAGAGTACTACATACAGTTGGAAATGCTTAAGACCCGGTACTA 264
QY 3282 AGCTGTTTCTTACTCAACCTGACCAAGATTTTCAGTAACACCAAAATTTATCTGCT 3341
Db |||||||
265 AGCTGTTTCTTACTCAACCTGACCAAGATTTTCAGTAACACCAAAATTTATCTGCT 324
QY 3342 TGAATGAATCATCTTTTCTCTCTGGAAGAACTTAACAACCAATGTTAGAGCTGT 3401
Db |||||||
325 AGAATGAATCATCTTTTCTCTCTGGAAGAACTTAACAACCAATGTTAGAGCTGT 384
QY 3402 TAACAAGCCACTTTCATCAGCAGCAAGCAATTCAGCAATATCATCAGATGAAGA 3461
Db |||||||
385 TATTAAGCCACTTTCATCAGCAGCAAGCAATTCAGCAATATCATCAGATGAAGA 444
QY 3462 TGTAAAGCAATGCAAGCAGCTCAATTCAGTCTCTCTGGAAGATTAAGGAGAGCT 3521
Db |||||||
445 TGTGAGCAACGCAAGCAGCTCCACCAAGCTCTCTGGAAGATTAAGGAGAGCT 504
QY 3522 TGATAGTCTGTAAGATGATCAGTGAAGAAATGAAGATTAACAAATGCTTCACTTGGC 3581
Db |||||||
505 TGATAGTCTGTAAGATGATCAGTGAAGAAATGAAGATTAACAAATGCTTCACTTGGC 563
QY 3582 GGGGAAAAAAGTGAAGAGAGAGAGTCTGATCTTGTAAAGTCTGATGGAAGAGC 3641
Db |||||||
564 AGGAAAAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 623
QY 3642 TAGAGGCAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3701
Db |||||||
3702 GACTAAGTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3760
QY 3762 GACTAAGTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3820
Db |||||||
744 GTACAGCTTTCAGACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 803
QY 3821 TATTAAGAAATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3880
Db |||||||
804 TCCTGAAATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 863

```

OY		3861	CAAAACCCCTTGG-TGSGAGGTACACCAAAGAAGGCCAACATGAATCCTTAATAA	3933
Dd		864	CAAAACCTTCCTTGNTGGGGGCACATCGAAGGAGAACCCACCATGAATTCCAAGAA	923
OY		3940 GG 3941	.	
Dd	-	924 AG 925	.	
RESULT 8	B0940978			
LOCUS		745 bp	mRNA	linear EST 18-OCT-2002
DEFINITION	AGENCOURT_10538461 NIH_MGC_128 Homo sapiens cDNA clone			
ACCESSION	B0940978			
VERSION	B0940978.1 GI:24129797			
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 745) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strusberg, Ph.D. Email: rgs@phs-remai.nhl.gov Tissue Procurement: NCI cDNA Library Preparation: Michael Brownstein Laboratory CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LUCM3020 row: n column: 16 High quality sequence stop: 553. Location/Qualifiers . . 745 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone IMAGE:6712336" /_tssname-type="mixed (pool of 40 RNAs)" /_lab_host="DH10B (TI-phage-resistant)" /_note="vector: pDNR-LIB; Site_1: SfiI (gagccattatggccc); Site_2: SfiI (ggcgcttgaggc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2% gland - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGTATACAACGAGCGGCGCATTTACGCGCGGG-3' and 5'-ATTCTAGAGCGCGGCGCGCACACTG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >2 kb size fraction (other fractions present in NIH_MGC_126 and NIH_MGC_127). Library created in the laboratory of T. Utsida, M.D., Ph.D. (NIMH, NIH). Note: This is a NIH_MGC library."			
BASE COUNT	252 A	142 C	152 G	199 T
ORIGIN				
Query Match	Best Local Similarity	99.6%	Pred. No. 1.5e-107;	Matches 698; Conservative 0; Mismatches 2; Indels 1; Gaps 1,
OY	1414 GAAGCGANGAANTGTTATTTACTTGTGATGCCACACGATTTAAAAAGCTGTGAACA	1473		
Dd	3 GAAGCGANGAANTGTTATTTACTTGTGATGCCACACGATTTAAAAAGCTGTGAACA	62		

OY		1474	TTGAAATGAATGTTGGAAATGTCAAATTCGTCTCCGACATCAAGTAAGGATTTGCTTGCAC	1533
Db		63	TTGAATGAAATGTGGAATGTCAAATTCGTCTCCGACATCAAGTAAGGATTTGCTTGCAC	122
OY		1534	TTGATTTAAGCAACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCCAAATGATGGTT	15933
Db		123	TTGATTTAAGCAACCCAAACAGATGCCAGTGTCAAGGCCATATTTTCCAAATGATGGTT	182
OY		1594	ATTACAAGAAATTTTACCTGATCCTGGTAAAGGCTCAGGATTTTCATGAAGAAATTCACACAG	16533
Db		183	ATTACAAGAAATTTTACCTGATCCTGGTAAAGGCTCAGGATTTTCATGAAGAAATTCACACAG	242
OY		1654	GTGTTTAGAAGATGATNGAAGAAATTAAGAAAGAGATTAGAAGTCTTTAGTCCCAACATGC	17133
Db		243	GTGTTTAGAAGATGATNGAAGAAATTAAGAAAGAGATTAGAAGTCTTTAGTCCCAACATGC	302
OY		1714	TCCTGCAAGCAGGCTGAAGTGTGTGGCTGAATAACTAGAAGTGGCGACACCCCAA	17733
Db		303	TCCTGCAAGCAGGCTGAAGTGTGTGGCTGAATAACTAGAAGTGGCGACACCCCAA	362
OY		1774	CAGCCTACAAATCCTTTCCTGGAATGATCAAGTTTCTCTTGGAGAGATATGACCTGTG	18333
Db		363	CAGCCTACAAATCCTTTCCTGGAATGATCAAGTTTCTCTTGGAGAGATATGACCTGTG	422
OY		1834	CACATAGTATACCGAATCTATCAGTGCCTTTTAAACAAGTAACAAATCAATATGATGGA	18933
Db		423	CACATAGTATACCGAATCTATCAGTGCCTTTTAAACAAGTAACAAATCAATATGATGGA	482
OY		1894	ACACAGATGATGAAGATGAGAGGTGTTCCAACTGATCAAGCCATCAGACGAGTCTTGA	19533
Db		483	ACACAGATGATGAAGATGAGAGGTGTTCCAACTGATCAAGCCATCAGACGAGTCTTGA	542
OY		1954	CTGCTTAAGTATCTCTCATTTTACACATGCCATCTCATTTTCATTCTGCTGGAACATTTGAA	20133
Db		543	CTGCTTAAGTATCTCTCATTTTACACATGCCATCTCATTTTCATTCTGCTGGAACATTTGAA	602
OY		2014	TCATTACGCGCTTCTGTAAGAAATGATGATGATAAAAGTAGACGAAGCTGCACACTCAAAAT	20733
Db		603	TCATTACGCGCTTCTGTAAGAAATGATGATGATAAAAGTAGACGAAGCTGCACACTCAAAAT	662
OY		2074	TTCAAAAAACACAGAGAG-CAAAATTTGAAGAGATTTTCCAC	2113
Db		663	TTCAAAAAACACAGAGAGCCAAAATTTGAAGAGATTTTCCAC	703
RESULT_9				
BM794657				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

FEATURES
source

Location/Qualifiers

1..692
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S22SN016n1-50-A08"
/sex="F"
/tissue_type="Asticles"
/cell_type="Lymphoblast-like"
/cell_line="SN0-16"
/lab_host="DH10B"
/clone_lib="S22SN016n1"
/note="Organ: Stomach; Vector: pTY73-Pac; Site:1: EcorI;
Site:2: NotI; The S22SN016 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldi, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SN0-16 culture. SN0-16 cell was obtained from
Korean Cell Line Bank (KCLB). SN0-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

BASE COUNT 264 a 136 c 148 g 144 t
ORIGIN

Query Match 16.2%; Score 677; DB 12; Length 692;
Best Local Similarity 99.3%; Pred. No. 4.9e-106;
Matches 680: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3186 TGAATACTGACACTGTGTGTGATGTTGCCATGAAATCATCATGTCAGAGTACTAC 3245
DB 1 TGAATACTGACACTGTGTGTGATGTTGCCATGAAATCATCATGTCAGAGTACTAC 60
QY 3246 ATACAGTTGGAATCTCTAAAGACCGGTACTACACTCGTTCCTCAGTCAACCTCGA 3305
DB 61 ATACAGTTGGAATCTCTAAAGACCGGTACTACACTCGTTCCTCAGTCAACCTCGA 120
QY 3306 CAGAGATTTGAGTAACACCAAAATATCTGCTCTCTGAAATGAATCATTTTTCAGTCC 3365
DB 121 CAGAGATTTGAGTAACACCAAAATATCTGCTCTCTGAAATGAATCATTTTTCAGTCC 180
QY 3366 TGGAAATCTTAAACACCAATGTTCTAGAGAGCTGTAAACAGCATTTCATCAGCAGG 3425
DB 181 TGGAAATCTTAAACACCAATGTTCTAGAGAGCTGTAAACAGCATTTCATCAGCAGG 240
QY 3426 CAAGCAATCTGAGCAACCAATCATCATGCAATGGAACCTGTAACAGTCAAGAGACGTC 3485
DB 241 CAAGCAATCTGAGCAACCAATCATCATGCAATGGAACCTGTAACAGTCAAGAGACGTC 300
QY 3486 AATTCAGAGCTCTCTGAGAGAAATTAAGGGAGGCTGATAGTTCTGAATGATGATCAG 3545
DB 301 AATTCAGAGCTCTCTGAGAGAAATTAAGGGAGGCTGATAGTTCTGAATGATGATCAG 360
QY 3546 TGAATGAGAGATTAACCAATGCTTCACTTTGCCGGGAAAAAAGTGACAGAGAGA 3605
DB 361 TGAATGAGAGATTAACCAATGCTTCACTTTGCCGGGAAAAAAGTGACAGAGAGA 420
QY 3606 CGACTCTGATCTTGAAGAGTCTGAATTTGAGAGAGGCTGAGAGCAAGAAAAACCCCGT 3665
DB 421 CGACTCTGATCTTGAAGAGTCTGAATTTGAGAGAGGCTGAGAGCAAGAAAAACCCCGT 480
QY 3666 CACAGAACAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATGATGAT 3725
DB 481 CACAGAACAGAGAGAAATTAAGTATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 3726 ACCTAAAGGAGTCAAGCAGAGTGGAAAAAGAGCCATAGGCTTCAGATCTGATGAACA 3785
DB 541 ACCTAAAGGAGTCAAGCAGAGTGGAAAAAGAGCCATAGGCTTCAGATCTGATGAACA 600
QY 3786 GCATGGCGCTGAGAGAAAGAGGCTCAAGAAGATATATTAAGAAATGAAGATGAACAGAA 3845
DB 601 GCATGGCGCTGAGAGAAAGAGGCTCAAGAAGATATATTAAGAAATGAAGATGAACAGAA 660
QY 3846 TAGTCGCGCAAAAAAGGTTAAAGA 3870

DB 661 TAGTCGCGCAAAAAAGGTTAAAGA 665

RESULT 10
CA324097 786 bp mRNA linear EST 26-NOV-2002
LOCUS
DEFINITION
UI-M-FY0-CCO-m-16-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
IMAGE: 6822425 5', mRNA sequence.
ACCESSION
CA324097
VERSION
CA324097.1 GI:24542195
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: egapbs.rem@nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES
source

Location/Qualifiers

1..786
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image: 6822425"
/tissue_type="Whole Brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FY0"
/note="Organ: Brain; Vector: pYX-Asc; Site:1: Ecor I;
Site:2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

BASE COUNT 256 a 171 c 165 g 191 t 3 others
ORIGIN

Query Match 16.1%; Score 671; DB 14; Length 786;
Best Local Similarity 90.7%; Pred. No. 5.3e-105;
Matches 713: Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 2794 CGTTACGCTTCCACTGATATGCAATCTGCTGCTTTGTGCAAAAGATCTGTA 2853
DB 1 CGCTTAGGCTTCCCTTGAGTACATGCGCATCTGCTTTGTGCCAAAGACCTGTG 60
QY 2854 AAGGAGAGAGAGCTCATGCTAGGCAATGTTGTGAAATAATTAATGTAAGCGGAG 2913
DB 61 AAGGAGAGAGAGAGCTCATGCTAGGCAATGTTGTGAAATAATTAATGTAAGCGGAG 120

OY	2914	TATCTGAAGCAGCATGCAAGCTGTATAGCAAAAATATATGTCTCTCTTACAGAGATGTT	2973
Db	121	TACCTGAAGCAGCATGCAAGCTGTATAGTCAAAAATATATGTCTCTTACAGAGATGTT	180
OY	2974	GTTCATATACAAATTCACCTTTTGGCAGATACCCAGATATATGTCAAAAGTACAGATAT	3033
Db	181	GTTCATATACAAATTCACCTTTTGGCAGATACCCAGATATATGTCAAAAGTACAGATAT	240
OY	3034	GACAACTTAAAGATGTTAAAGATGTCCTTTGGTTTGTCTCGAAATATTAATGGCTAA	3093
Db	241	GAAACAACCTTAAAGATGTCGAAAAGATGCTTTGGTTTGTCTCGGAATATTCATGGCTAA	300
OY	3094	AATGAAATTAACAGTACAGCTTTTATATCGAAAAGTGGTAGAAAATATATTAACAAACAA	3153
Db	301	AATGAAATTAACAGTACAGCTTTTATATCGAAAAGTGGTAGAAAATATATTAACAAACAA	360
OY	3154	GATGCCCAAGCAGCAGATGATGCAAAAAATGAATGAAAACTGTACACTGTGTGTGATGT	3213
Db	361	GATGCTCAAGGACCAAGATGATACAAAAATGAATGAAAAATGTACACCTGTGTGATGTT	420
OY	3214	GCCATGAATATCATCATGTCTCAAGAGTACTATACATGATTTGGAATCTCTCAAAAGACCG	3273
Db	421	GCCATGAATATCATCATGTCTCAAGAGCACCACGTACAGCTGTGAGTCTCTCAAGAGACCC	480
OY	3274	GTACTACAGCTGTTTCTTCACTCACTCAACCTGACAGAAATTTCACTAACACCAAAATTT	3333
Db	481	GTCGTGCAGCTCGTGGTTTTTACACCAGCTTACAGAAATTTAGTAAACCCAAAAATATAC	540
OY	3334	CTGCCTCCCTGAAATGAATATCATTTTTCACCTGCTGGAAAACCTTAACCAACCAATGTTCTA	3393
Db	541	CTGCCTCCAGAAATGAATATCATTTTTCACCTGCTGGAAAACCTTAACCAACCAATGTTCTC	600
OY	3394	GGAGCTGTTAACAAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATTCATCAGCA	3453
Db	601	GGAGCAGTTTAAAGGCCACTTTCATCAGCAGGCAACAGCTCTCAGACCAATTCATCAGCA	660
OY	3454	ATGGAATCTGTAAGCAATGCAAGCAGCAGCTCCAATCCAACTCTCTGGAGAAATTAAG	3513
Db	661	ATGGAATCTGAGAACGCAAGCAGCAGCTCCCAACCCAACTCTCTCGAAGGATCAAG	720
OY	3514	GGAGGCTGTGTAAGTCTGGAATGGATCACAGTGAATAATGAAGATATACACATGCTCTCA	3573
Db	721	GGAGGCTGTGTAAGTCTGGAATGGATCACAGTGAATAATGAAGATATATACATGCTCTCA	780
OY	3574	CCTTTG 3579	
Db	781	CCTTTG 786	
RESULT 11			
BG256731			
LOCUS	603270939p1	NIH_MGC_92 Homo sapiens CDNA	linear EST 13-FEB-2001
DEFINITION	mRNA sequence.		
ACCESSION	BG256731		
VERSION	BG256731.1	GI:12766547	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 734)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs.femail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at:		

[illegible]

Db 720 CAAA 724

|||||

RESULT 12

LOCUS CB312535

DEFINITION AGENCOURT_11835698 NICHD_Rh_Ov1 Macaca mulatta cDNA clone

IMAGE:6892555 5', mRNA sequence.

ACCESSION CB312535

VERSION CB312535.1 GI:28835251

KEYWORDS EST.

SOURCE Macaca mulatta (rhesus monkey)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.

REFERENCE 1 (bases 1 to 771)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Eliot Spindel
cDNA Library Preparation: Clontech
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://lml.nhl.gov>
Plate: LICM3155 row: 9 column: 18
High quality sequence stop: 563.
Location/Qualifiers

FEATURES

source 1..771

/organism="Macaca mulatta"

/mol_type="mRNA"

/db_xref="taxon:9544"

/clone="IMAGE:6892555"

/rissue_type="Ovary"

/lab_host="NIDH (phage-resistant)"

/clone_lib="NIDH_Rh_Ov1"

/note="Organ: ovary; Vector: pDNR-LIB; Site1: Sfi I; Site2: Sfi I; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.0-4.0 kb. Tissue pooled from pre-pubertal, post pubertal sn menopausal monkeys. Constructed by Clontech. Note: this is a NICHD library."

BASE COUNT 291 a 144 c 159 g 177 t

ORIGIN

Query Match 15.8%; Score 660.6; DB 14; Length 771;
Best Local Similarity 96.0%; Pred. No. 3.2e-103;
Matches 744; Conservative 0; Mismatches 19; Indels 12; Gaps 6;

QY 3014 ATGTCAAGTACAGATATTGAACAACCTTAAGATGTGAAGATGCTTTGGTTGTTTC 3073

Db 3 ATGTCAAGTACAGATATTGAACAACCTTAAGATGTGAAGATGCTTTGGTTGTTTC 62

QY 3074 TCGAATATTAATGCTTAATAATGAATAATACAGTCACGCTTTTATCAGAAAGATGTAG 3133

Db 63 TCGAATATTAATGCTTAATAATGAATAATACAGTCACGCTTTTATCAGAAAGATGTAG 122

QY 3134 AAAATATTAACAACAAGATGCCCAAGACCGATGATGCAAAAATGAATGAAGAAC 3193

Db 123 AAAATATTAACAACAAGATGCCCAAGACCGATGATGCAAAAATGAAGAAC 182

QY 3194 TGTACACTGTGTGTGATGTTGCCATGATATCATCTGTCAAGAGTACTACATACAGTT 3253

Db 183 TGTATACCGTGTGTGATGTTGCCATGATATCATCTGTCAAGAGTACTACATACAGTT 242

QY 3254 TGGATCTCTCAAGACCCGCTGACTACGAGCTGTTTCTTCACTCACTGACAGCAAT 3313

Db 243 TGGATCTCTCAAGACCCGCTGACTACGAGCTGTTTCTTCACTCACTGACAGCAAT 302

QY 3314 TCAGTACACCAAAAATTTATCTGCGCTCGAAGTGAATCATTTTCACTCTCGGAAAC 3373

Db 303 TCAGTACACCAAAAATTTATCTGCGCTCGAAGTGAATCATTTTCACTCTCGGAAAC 362

QY 3374 CTAAAAACAACCAATGTTCTAGAGAGCTGTTAACAGCCACTTTCATCAGGCAAGCAT 3433

Db 363 CTAAAAACAACCAATGTTCTAGAGAGCTGTTAACAGCCGCTTTCATCAGGCAAGCAT 422

QY 3434 CTCAGACCAATTCATCAGATGGAATGAAACTGTAAAGCAATGCAAGCAGCTCAATCCA 3493

Db 423 CTCAGACCAATTCATCAGATGGAATGAAACTGTAAAGCAATGCAAGCAGCTCAATCCA 482

QY 3494 GCTCTCCTGGAAGAAATTAAGGGAGGCTGATAGTTCTGAATGATCAGTGAAGAAATG 3553

Db 483 GCTCTCCTGGAAGAAATTAAGGGAGGCTGATAGTTCTGAATGATCAGTGAAGAAATG 542

QY 3554 AAGATTACCAATGCTCTCACCCTTTCGCGGGAAAAAAGTGAACAAGAGACGACTCTG 3613

Db 543 AAGATTACCAATGCTCTCACCCTTTCGCGGGAAAAAAGTGAACAAGAGACGACTCTG 602

QY 3614 ATCTTGAAGTCTGTGATTTGAGAGAGCCCTAGAGCAGAAAAAAGCCGCTCAGCAAC 3673

Db 603 ATCTT-----TCTGAATTGAGAGAGCCCTAGAGCAGAAAAAAGCCGCTCAGCAAC 656

QY 3674 AGGAG-GAGAAATTTGCTATGATGATGAC-TTGACTAAGTTGTTAGACAGGAAC-AGAAACCTA 3730

Db 657 AGGAGGAGAAATTTGCTATGATGATGACTTGTGACTAAGTTGTTAGACAGGAACAGAAACCTA 716

QY 3731 AAGGCACTGAG-CGAAGTCGAAAGA--GGCCATACGCTTCAGATCTGATGA 3782

Db 717 AAGGCACTGAGCGAGGTGGAAGAAAGCCATACGGCTTTCGATTTGATGA 771

RESULT 13

LOCUS CA405800

DEFINITION 1001949 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens

CDNA 5', mRNA sequence.

ACCESSION CA405800

VERSION CA405800.1 GI:24770671

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS Yang, R.-Z., Shuldiner, A. and Gong, D.-W.

TITLE EST analysis of human adipose gene expression

JOURNAL Unpublished

COMMENT Contact: Gong Da-Wei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St., BH497, Baltimore, MD 21201, USA
Tel: 410 706 1672
Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu
PCR Primers
FORWARD: CTCGGGAGCGCGCCATGCTGTGCT
BACKWARD: AATAGACTCACTATGAGCGCAATTCG
Seq primer: GTTGGTACCCGGGAATTC.
Location/Qualifiers

FEATURES

source 1..695

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/sex="Male and Female"

/rissue_type="Adipose"

/clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"

/note="Vector: lambdaTriplex"

BASE COUNT 222 a 154 c 128 g 191 t

ORIGIN

Query Match	15.7%	Score 657.2	DB 14	Length 695
Best Local Similarity	99.1%	Pred. No. 1.2e-102		
Matches 692	Conservative 0	Mismatches 3	Indels 3	Gaps 3
QY	1680	AAAGCGATTGAGAGTCTTGTAGTCCAAATGCTCCGACAGAGGCTGAAGTTGTG	1739	
DB	1	AAAGCGATTGAGAGTCTTGTAGTCCAAATGCTCCGACAGAGGCTGAAGTTGTG	60	
QY	1740	GCCTGAAATPACTAAGAGTTGGCAACCCCAAGCCCTCAAAATCCTTCTGGAAT	1799	
DB	61	GCCTGAAATPACTAAGAGTTGGCAACCCCAAGCCCTCAAAATCCTTCTGGAAT	120	
QY	1800	GATCAAGTTCTCTTGGAGAGATAGACCTGTGCACATGATACCAATCTATCACTGC	1859	
DB	121	GATCAAGTTCTCTTGGAGAGATAGACCTGTGCACATGATACCAATCTATCACTGC	180	
QY	1860	TCTTATTAACAAGTGAACAATCAATAGATGAGACAGAGATGATGAATGAGGTGT	1919	
DB	181	TCTTATTAACAAGTGAACAATCAATAGATGAGACAGAGATGATGAATGAGGTGT	240	
QY	1920	TCCACTGATCAAGCCATCAGACAGGCTTGAAGCTTGAAGTACTTCAATTTACACA	1979	
DB	241	TCCACTGATCAAGCCATCAGACAGGCTTGAAGCTTGAAGTACTTCAATTTACACA	300	
QY	1980	TCCATCTCATTTCTGCTGCTGAACATTTGATCTAGCTGCTGCTGAACATTTG	2039	
DB	301	TCCATCTCATTTCTGCTGCTGAACATTTGATCTAGCTGCTGCTGAACATTTG	360	
QY	2040	TGATGAAAAAGTAGACAGACAGCTGCTGCTGAACATTTGAAAAAGACAGAAATTTGA	2099	
DB	361	TGATGAAAAAGTAGACAGACAGCTGCTGCTGAACATTTGAAAAAGACAGAAATTTGA	420	
QY	2100	AGAGGATTTTCCACACATCAGATCAGCTTCTGCTTCTTACATCAAAATCTAAAA	2159	
DB	421	AGAGGATTTTCCACACATCAGATCAGCTTCTGCTTCTTACATCAAAATCTAAAA	480	
QY	2160	AGAGACCCCGGCTCAACCAATATGATGATGATGATGATGATGATGATGATGATG	2219	
DB	481	AGAGACCCCGGCTCAACCAATATGATGATGATGATGATGATGATGATGATGATG	540	
QY	2220	AGAGACCCCGGCTCAACCAATATGATGATGATGATGATGATGATGATGATGATG	2279	
DB	541	AGAGACCCCGGCTCAACCAATATGATGATGATGATGATGATGATGATGATGATG	600	
QY	2280	GGAACATCTCATTAACACATGCTTACTATTTGCTGATATTTCTCTTGCACCTATCA	2339	
DB	601	GGAACATCTCATTAACACATGCTTACTATTTGCTGATATTTCTCTTGCACCTATCA	658	
QY	2340	ATTGCTGCTCTGGAATCTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2377	
DB	659	ATTGCTGCTCTGGAATCTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	695	
RESULT 14				
LOCUS	BU703488	786 bp	musculus	EST 09-OCT-2002
DEFINITION	UI-M-F00-b20-g-20-0-UI.r1 NIH-BMAP_F00		musculus	CDNA clone
ACCESSION	BU703488			
VERSION	BU703488.1	GI:23629371		
KEYWORDS	EST			
SOURCE	musculus (house mouse)			
ORGANISM	musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	NIH-MGC http://mgc.ncl.nih.gov/			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgaabs-r@mail.nih.gov			
	Tissue Procurement: Dr. Jim Lin, University of Iowa			
	CDNA Library preparation: Dr. M. Bento Soares, University of Iowa			

FEATURES		Seq primer: PYX-5.	
source	Location/Qualifiers		
	1. 786		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6"		
	/db_xref="taxon:10090"		
	/clone="IMAGE: 6405163"		
	/tissue_type="whole brain"		
	/dev_stage="embryo 12.5dpc"		
	/lab_host="DH10B (T1 phage resistant)"		
	/clone_id="NIH-BMAP_F00"		
	/note="Organ: Brain; Vector: pyx-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction. ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGAGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."		
BASE COUNT	248 a 143 c 152 g 242 t 1 others		
ORIGIN			
Query Match	15.7%; Score 653.8; DB 13; Length 786;		
Best Local Similarity	91.1%; Pred. No. 4.7e-102;		
Matches 694	Conservative 0; Mismatches 68; Indels 0; Gaps 0		
QY	19 AGGACCATGATGAAAAATTAATATCCCTGGGCTCAAGAAATATCATATAAATA	78	
DB	25 AGGACCAACGATGGGAAAAATTAATTAACCTCCTGGAGTCMAAGAAATCTCAATATAAAC	84	
QY	79 TCTAAGAGGAGATGTGAGACGATTAAGATGGTGTGAACCTTTATGATATGAC	138	
DB	85 TCTAAGAGGAGATGTGAGAGCGGTTAAGATGGTGTGAACCTTTATGATATGAC	144	
QY	139 CAGGACTCTGAAGAAAAAGAGCTTATTTAACTTACTTACTTCTGCTAGAT	198	
DB	145 CAGGACTCTGAAGAAAAAGAGACTTTATCTAAACCTATGCTTTACATCTTCTGCTGAC	204	
QY	199 TTTTTCCTCAAGCATCCCTGGTAAAGATGTGCTTACTGCTAGCTCTGCTGAT	258	
DB	205 TTTCTCCCAAGCATCCCTGATTAAGATGTGCTTACTGCTAGCTCTGCTGAT	264	
QY	259 ATTTTCAGATTTATGCTCTGAGAGCTCTTACACATCCCTGATTAAGATATA	318	
DB	265 ATTTTCAGATTTATGCTCTGAGAGCTCTTACACATCTCCCGATTAAGATATA	324	
QY	319 TTTATGTTTATTAACACAGCTTGAAGGGCTAGAGATCAAAAGAGCCCAATTAAT	378	
DB	325 TTTATGTTTATTAACACAGCACTAAGAGGACTAAGAAATCAAAAGAGCCCTCAATTTAAT	384	
QY	379 AGGATTTTATTTACTTGAACATGCTGGGTCAAGCATATTAATGCTTTGAG	438	
DB	385 AGGATTTTATTTACTTGAACATGCTGGGTCAAGCATATTAATGCTTTGAG	444	
QY	439 TTAGAAGATAGCAATGAATTTTCCAGAGCTATACAGAACTTATTTGATTAATAC	498	
DB	445 TTAGAAGATAGCAATGAATTTTCTTACTCAACTTACAGAACTTATTTGATTAATAC	504	

[illegible]

Query Match	Best Local Match	Local Similarity	Score	DB	Length	779;
Matches	699;	Conservative	0;	Pred	No. 8.6e-100;	
			0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;
Best Local Match	89.6%;	Pred	No. 8.6e-100;			
Matches	699;	Conservative	0;	Mismatches	80;	Indels 1;
				Gaps	1;	
BASE COUNT	241 a	166 c	174 g	197 t	1 others	
ORIGIN						
Query Match	15.4%;	Score	640.6;	DB	14;	Length 779;

THIS PAGE BLANK (USPTO)

QY 2507 TACTGGAATGAAAAATATATCAGATAATCAGAACTTCTACCTTAAGATTGCTAACAA 2566
 Db 203 TGTGGGTATGATAAACAACAGCTCTAATCTGCCAATTCACACCTTGCTGTTATATACAG 262
 QY 2567 CAATATTGCATAGATGATGAGAACTTGACAGAAACAGGGGAAAAATTATGTAACCAATANGT 2626
 Db 263 CGATGTGGTTAGTAGAGGGTGCCTGACAGAGCAAAAAGAGATCACTGTAATCTGATTAAGT 322
 QY 2627 CAGCTCTGAGACTTGTGCTGGAGTGTCTATTTGTAAGCTGGCACACAAAGACCTGTAC 2686
 Db 323 CTCCTCTCCGATATAGCTGTCTGTAGTGCCTATATGAACTTGTCTAGAACCTTGTAC 382
 QY 2687 ATGAATATCATCATTTAGAACAAATATCAGTATGTGATTAATGATCAACATGATGAT 2746
 Db 383 ATGAATATTTATACCCAGAACAGTTTCAGTCTGTGCACTTGTATTAATGATGATGAT 442
 QY 2747 ATCAAGTAAAGCAAGTGTGTTGCCAGAACTTCACAAAGGCTTCCCGTTACGGCTTC 2806
 Db 443 ACCAAGTAAAGCAATATTTGCTCTCAAGAGCATTAAGGCACTTGTGAAGTTACTGCTCC 502
 QY 2807 CACTTGAATATATGCAATCTGTGCTGCTTGTGCAAAAGATCCTGTAAGAGAGAGAGAG 2866
 Db 503 CATGGAGATATATGGCAATCTTGTGCTGTGTGCCAAAGATCCTGTGAAGAGAGAGAGAG 562
 QY 2867 CTCAATGCTAGCAATGTTTGTGTAATAATATTAATGTAAGCGGAGATATCTGAAGACAG 2926
 Db 563 CACAGCGACGACAAATGTTTACTGAAAAATATCAGATATACGACAGGAAATACATTTAAGCAGA 622
 QY 2927 ATGACAGTGTAGTAAATAATTTATGCTCTTCAACGAGATATGCTTCCATATACAA 2986
 Db 623 ATCTATATGCTACTGAGAAATTTATATCTACTGTGCTGAAATATGTAATCTCATACATGA 682
 QY 2987 TTCACCTTTTGGACATGACCCAGATATATGTCAAAGTATCAGAGATATGACAACTTAAG 3046
 Db 683 TTCACCTCTGACCCATGATCCAGATTTTACAAGATTCACAAAGATTTATCATGCTGCTG 742
 QY 3047 ATGTTAAGATATGCTTGTGTTGTGTTGTAATAATTTATGCTTAAATGAAATTAACA 3106
 Db 743 ATATCAAAAGATGCTATGCTTATGCTTGAAGTTTATGACAAAGATGAAAAACAATA 802
 QY 3107 GTCACGCTTTATCAGAAAGATGTAAGAAATATTAACAACAAAGACCCCAAGAGAC 3166
 Db 803 GCCATGCTTTATGAGAGAGATGGCAGAGAAATCAATGTTAACAGAGATGCCAGTCTC 862
 QY 3167 CAGATGATGCAAAAAATGAAATGAAATGTAACATGCTGTGTGATGTTGCCATGATATCA 3226
 Db 863 CAGATGATGCAAAAGCAAAATGAAATGTAACATGATGATGATGCTGCTGCTGTTA 922
 QY 3227 TCATGTCAAAGATATCATACATGTTGGAATCTCTTAAGACCCGCTACTACAGCTC 3286
 Db 923 TAAATATGTAAGAGCTTTGTGCA - ATGCAAGATTCACCAAGAGACCCGCTCCAAATGAA 980
 QY 3287 GTTCTCTACTCAACCTGACAAAGATTCAGTAACACCAAAATTAATGCTGCTCTCTGAAA 3346
 Db 981 ATTTTACAC - - - ACCTGAAAAGAGCTTCTGTATAGATTAAGAGTTATTTTCAGAGAGA 1037
 QY 3347 TGAATCATTTTTCACCTGCTGAAAAACCTTAAACCAACCAATGTTCTAGAGCTGTTACA 3406
 Db 1038 CAAAGATACCTTCTGTAAAGAGAAAGCCGCTGAGAGTACTAGTGTGAGTAATA 1097
 QY 3407 AGCCACTTTCATGACGAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGGAATGTAA 3466
 Db 1098 AGCCTTATACGCAAGGGAAGAAACCTTATGTTAGAGCACTGGCACTAGACTGTGAAA 1157
 QY 3467 GCAATGCAAGAGAGCTCAATCCAACTCTCTCGAAGAAATTAAGGAGGCTGTGATA 3526
 Db 1158 GCAATATTAATTAATTCAGAGCTGAACCTTCACCGGAAATTCATCAAGGGAACAGA 1217
 QY 3527 GTTCTGA 3533
 Db 1218 GTTCAGA 1224

RESULT 2
 US-10-040-739-107
 ; Sequence 107, Application US/10040739
 ; Publication No. US20020173635A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; McCoy, John
 ; LaVallie, Edward
 ; Racie, Lisa
 ; Merberg, David
 ; Spaulding, Vikki
 ; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
 ; NUMBER OF SEQUENCES: 1519
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/040,739
 ; FILING DATE: 07-Jan-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/036,520
 ; FILING DATE: 03-JUN-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brown, Scott A.
 ; REGISTRATION NUMBER: 32,724
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 498-8224
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 439 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 107:
 US-10-040-739-107
 Query Match 9.9%; Score 411.4; DB 13; Length 439;
 Best Local Similarity 98.6%; Pred. No. 1.5e-91;
 Matches 415; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 3216 CATGAATATCATCATGTGTAAGAGATGATACATGTTGGAATCTCTTAAGACCCGCT 3275
 Db 13 CATGCCCTCAATCAATGCTCAAGAGATGATACATGTTGGAATCTCTTAAGACCCGCT 72
 QY 3276 ACTACACGCTGTTTCTTCACTCAACCGACAGCAATTTCACTAACCACCAAAATTAATCT 3335
 Db 73 ACTACACGCTGTTTCTTCACTCAACCGACAGCAATTTCACTAACCACCAAAATTAATCT 132
 QY 3336 GCTCTCGAATATGAATATTTTCACTCTGGAAGAAACCTTAAACCAACCAATGTTCTAGG 3395
 Db 133 GCTCTCGAATATGAATATTTTCACTCTGGAAGAAACCTTAAACCAACCAATGTTCTAGG 192
 QY 3396 AGCTGTTAACAAGCCACTTTATCAAGAGGCAAGCAATCTGAGACCAATATCAAGAT 3455
 Db 193 AGCTGTTAACAAGCCACTTTATCAAGAGGCAAGCAATCTGAGACCAATATCAAGAT 252
 QY 3456 GGAACCTTAAGCAATGCAAGCAGAGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGG 3515
 Db 253 GGAACCTTAAGCAATGCAAGCAGAGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGG 312

Qy	3253	TTGGAATCTCTCTAAAGACCCGGTACTACACCTGTTTCTTCACTCAACCGACGACAGAT	3312
Db	362	GCACATTCACCCAAAGAGACCAGTCTCCCAATGAAATTTTTTTTACACACCTGAAAGAGAC	421
Qy	3313	TTTCAGTAAACCAAAAATTTATCTCCCTCCCTGAAATGAAATCATTTTTCACCTCTGAAAA	3372
Db	422	TTCTGTAAAGCAATAGAGTTTATATTTCAGAAAGACACAGAGTACTGTGTTAAAGGAAAG	481
Qy	3373	CCCTAAACACCAATGTTTCTTAGAGAGCTGTTTAAACAGCCACTTTTATCATCAGCAGCAACAA	3432
Db	482	CCAAAGCCCTGCTGGAGTACTGAGTGCAGTAATTAAGCCTTTATCAGCAACCGGAAAGAAA	541
Qy	3433	TTCTAGACCAAAATATATCAAGAAATGAAACGTAAAGCAATGCAAGCAGAGCTCAAAATCCA	3492
Db	542	CCCTATGTTTGAAGACACTGGCAGCTAGAGCTGGAAGCAATATTATGTAAATTTCAAGACTG	601
Qy	3493	AGCTCTCCTGGAAAGATAAAGGGAGGAGCTTGATAGTTCTGA	3533
Db	602	AACCTTCACACCGGAATTCATCAAGGAAACAGAGTTTCAGA	642
RESULT 4			
US-09-918-995-3514/c			
: Sequence 3514, Application US/09918995			
: Publication No. US20030073623A1			
: GENERAL INFORMATION:			
: APPLICANT: Hyseq, Inc.			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED			
: TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES			
: FILE REFERENCE: 20411-756			
: CURRENT APPLICATION NUMBER: US/09/918,995			
: CURRENT FILING DATE: 2001-07-30			
: PRIOR APPLICATION NUMBER: US/09/235,076			
: PRIOR FILING DATE: 1999-01-20			
: NUMBER OF SEQ ID NOS: 38054			
: SOFTWARE: FastSeq for Windows Version 3.0			
: SEQ ID NO 3514			
: LENGTH: 471			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-918-995-3514			
Query Match			
Best Local Similarity 5.2%; Score 217.6; DB 11; Length 471;			
Matches 298; Conservative 0; Mismatches 134; Indels 0; Gaps 0;			
Qy	1309	AAAGACAAATTTGCATCATATATATTATCAAAATAGTATGATGATCGACTACTTGTGAA	1368
Db	459	AAAGGCAAACTCTCGCATTTATTATTCAGAACGACATTTGACGACAAACACTGTTGTAGAG	400
Qy	1369	CGGATCTTTGGCTCAATACATGTTCTTCACAAATTTAGAAACTACAGAACGAGTAATGC	1428
Db	399	AAATCTTTGCTCAGTATCTTGTCCCAACCTGGAAACAGAAAGAGATGAATGATGC	340
Qy	1429	TTATATACTTGTATGTCACACGTGATTTAAATGCTCTGAAAGCATGTAATGAAATGTGG	1488
Db	339	TTATATACTTATATGCTAGCTTTGGATCCAAATGCTTAATAAGCTCTCAACGAAATGTGG	280
Qy	1489	AAATGTCMAAATCTGCTCCGACATCAATCAATGAAGGATTTGCTTGACTTGATTAGCAACCC	1548
Db	279	AAGTGTCAGAACATGCTTCGGAGCGCATGACGGAACATTAATGGATTTTGCAACAGCAGCT	220
Qy	1549	AAAACAGATGCGCACTGTCAAGGCCATATTTTCAAAAGTAGTGATGTTATTACAGAAATTTA	1608
Db	219	ACATCAAGAGCTAACTGTTCTGCTATGTTGGAAACCTGATGACCATAGCAAAAGAAATTTG	160
Qy	1609	CCGATCTCTGCTGAAGGCTCAGAGATTTCATGAAGAAATTCACACAGGTGTTAGAGAGATGAT	1668
Db	159	CTGACCCCGGGAAGCACAAGATTTTGTGAGAAATTTTACACAGGTTTCTCGCGGATGAT	100
Qy	1669	GAGAAATAAGAAAGCATTAGAGTACTGTTTAGTCCAAATGCTCTCTGCAAGCAGCT	1728

DB 99 GAGAACTCGCTCACTTGAGTATTATTAATACCCCAACCTGTTTGCAGAAAGACA 40
QY 1729 GAAGTTGTGTG 1740
DB 39 GATATTGTGTG 28

RESULT 5

US-09-879-536-682/c
; Sequence 682, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Carino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 682
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-879-536-682

Query Match
Best Local Similarity 98.1%; Score 203.6; DB 10; Length 530;
Matches 206; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 104 TAAAGATGTTGTGAAAACTTTATGATATGACAGACACTGTAAGAAAAAGAGC 163
DB 218 TATGATGCTGTGAAAACTTTATGATATGACAGACACTGTAAGAAAAAGAGC 159
QY 164 TTTATTAACTTACTTACATCTTGCTCAGATTTTCTCAAGATCCTGTAAAG 223
DB 158 TTTATTAACTTACTTACATCTTGCTCAGATTTTCTCAAGATCCTGTAAAG 99
QY 224 ATGTTGCTTACTGTAGCTGCTGCTGCTGCTGATATTTTCAAGATTTATGCTCTGAAG 283
DB 98 ATGTTGCTTACTGTAGCTGCTGCTGCTGCTGATATTTTCAAGATTTATGCTCTGAAG 39
QY 284 CTCCTTACACATCCCTGATTAAG 313
DB 38 CTCCTTACACATCCCTGATTAAG 9

RESULT 6

US-09-918-995-33817
; Sequence 33817, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 33817
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33817

Query Match
Best Local Similarity 100.0%; Score 164; DB 11; Length 433;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4009 ACGGACGAGAGTCCAAAGCAACAGCAGCAGTGTCAAGAGAGCACAGCAGAGCA 4068
DB 3 ACGGACGAGAGTCCAAAGCAACAGCAGCAGTGTCAAGAGAGCACAGCAGAGCA 62
QY 4069 GAATCTCGAATCTAGTCAATTTGATTCACACAGTCACACACAGAGAGAGAGCA 4128
DB 63 GAATCTCGAATCTAGTCAATTTGATTCACACAGTCACACACAGAGAGAGAGCA 122
QY 4129 AGACCATCAAAAACGCCATCACCATCACACCAAAAAAATGT 4172
DB 123 AGACCATCAAAAACGCCATCACCATCACACCAAAAAAATGT 166

RESULT 7

US-09-918-995-20720
; Sequence 20720, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 20720
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-20720

Query Match
Best Local Similarity 74.7%; Score 96.4; DB 11; Length 413;
Matches 121; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 107 AGATGTTGTGAAAACTTTATGATATGACAGACACTGTAAGAAAAAGAGCTTT 166
DB 252 AGATGTTGTGAAAACTTTATGATATGACAGACACTGTAAGAAAAAGAGCTTT 311
QY 167 ATTAAACCTTACTTACATCTTGCTCAGATTTTCTCAAGATCCTGTAAAGATG 226
DB 312 ATCTCCACATAGCCTTGCATCTTGATCTGATTTCTCAGAGAACCCCAATAAGATG 371
QY 227 TTCGCTTACTGTAGCTGCTGCTGCTGCTGATTTTCAAG 268
DB 372 TCGCTTCTCTTGTACATGTTGTTGGCTGATTTCTTCGA 413

RESULT 8

US-09-728-445-172
; Sequence 172, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:

; APPLICANT: Friedlich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
Animals

```
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172
; LENGTH: 331
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(331)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-172

Query Match          2.1%; Score 87.4; DB 10; Length 331;
Best Local Similarity 85.1%; Pred. No. 2.7e-11;
Matches 97; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 190 GCTTCAGATTTTCTCAAGACCTGTAAGAGTTGCTTACTGTAGCCTGCTGC 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 216 GATTCGACCTCTCTCCANGCAGCCTGATTAAGATGTCTTACTGTGCTGCTGC 275

QY 250 CTTCGTGATTTTTCAGGATTTTNGCTTCGAGCTCCTTACACATCCCTGAT 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 276 CTTCGTGATTTTTCAGGATTTTNGCTTCGAGCTCCTTACACGTGAGAT 329

RESULT 9
US-10-027-632-105871
; Sequence 105871, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105871
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-105871

Query Match          1.8%; Score 76.8; DB 13; Length 642;
Best Local Similarity 90.0%; Pred. No. 1.8e-08;
Matches 81; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
```

```
RESULT 10
US-10-027-632-19016/c
; Sequence 19016, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19016
; LENGTH: 739
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-19016

Query Match          1.8%; Score 76.8; DB 13; Length 739;
Best Local Similarity 90.0%; Pred. No. 1.9e-08;
Matches 81; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2388 TCTTCATGATGATGCGGTTCCAGGAGAAAGACACAACTTGGGTCCAGATGA 2447
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 272 TTTTCCTTTATTTTATTTTATTTAGCTTCAGGAGAAAGACAACTTGGGTCCAGATGA 213

QY 2448 AGAAGTATCTCTGAGACATGTCATAAT 2477
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 212 AGAAGTATCTCTGAGACATGTCATAAGT 183

RESULT 11
US-10-311-455-1931/c
; Sequence 1931, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10033529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1931
; LENGTH: 14006
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```


[illegible]

OY	3611	CTGATCTTGTAAGGTCTGCAATTGGAGAAGCCTAAGAGCGAGAAAAAACGCCCTCACAG	3670
Dd	567	AAATCAAAAAAAAA	508
OY	3671	AACAGAGAGAAATTTAGTGTGTGATGCCTTACTGAAGTTGGTCAGAGAACAGAACTTA	3730
Dd	507	AAAAAAAAAATTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAA	448
OY	3731	AAGCAGTCACGGAATCGGAAAAGAGCCCTTACGGCTTCAGATCTGATGACAGCACT	3790
Dd	447	AAAAATAAGCAAAAAAAAAACA	388
OY	3791	GGCTTGAGAAAGAGGCTCAAAGAGATTAATTAGAAAAATGAGATGAACAGAAATGTC	3850
Dd	387	AATTTAAAC	328
OY	3851	CGCCAAAAAAGGTAAAGAGAGCCGACCACCAAACTCTTGTGGAGTTACACCAAAAG	3910
Dd	327	AAATTTTTAATTTAAAAAAAAAAAAACGAAAAAAAAAAAAAAAAAAAAACCAACGACAAAAAAA	268
OY	3911	AAGAGCCAACAATGAAAACTCTAAAAAAAAAGAACCAAAAAAATCTGGACCTCCACAC	3970
Dd	267	AAAAACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAACACCAAAATTTAAACGAAAAAAAAATCAAAAA	208
OY	3971	CAGAGGAGGAGGAAGAAAGAAAGACAAACTGGAAATACGGAACAGAAAGTCCAAAAAGCA	4030
Dd	207	AATCTTTAAATTTAAA	148
OY	4031	AACAGACCCGAGTGTCAAGGAGACACAGCAGAGACAGAAATCTCTGAAATCTTAGTCAA	4090
Dd	147	AAAAAAAAAAAAAAAAAATTTAAAAAAAAAAAAAAAAAATTTAAAAAATAATTTAAAAATTTAAA	88

```

      Db          87 AAAAAAAAAAAAAAAAAACAAATAAAAAAAAAATAAAAAAAA 37

RESULT 15
US-10-239-676-52/C
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

```


[illegible]

```

RESULT 4
US-09-268-992-7/c
Sequence 7, Application US/092686992
Patent No. 6342351
GENERAL INFORMATION:
APPLICANT: Chen, H.
APPLICANT: Fremler, N.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
FILE REFERENCE: 7853-138
CURRENT APPLICATION NUMBER: US/09/268, 992
CURRENT FILING DATE: 1999-03-16
EARLIER APPLICATION NUMBER: 09/236,134
EARLIER FILING DATE: 1999-01-22
EARLIER APPLICATION NUMBER: 60/106,056
EARLIER FILING DATE: 1998-10-28
EARLIER APPLICATION NUMBER: 60/088,312
EARLIER FILING DATE: 1998-06-05
EARLIER APPLICATION NUMBER: 60/078,044
EARLIER FILING DATE: 1998-03-16
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 72604
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: all n positions
OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

```

	Query Match	1.28;	Score 51.4;	DB 4;	Length 72604;
	Best Local Similarity	46.18;	Pred. No. 0.0076;		
	Matches	172;	Conservative	0;	Mismatches 201; Indels 0; Gaps 0;
QY	3693	GGATGACTGTGACTAAGTTGGTACAGGAACAGAAACCTAAAGCAGTCAGCGAATCGGAA			3752
DB	7381	GGATACCTGTAGGCCCAAGGGGTTCTAGCTGACCTGACACATCATGTGAGAACTGGTGGAA			7322
QY	3753	AAGTGGCCATTCGGCTTCAGAAATCGATGATGAAACACAGTCGGCTTAGGAAAAAGAGGCTTCAA			3812
DB	7321	AGAAAGCAAGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA			7262
QY	3813	AGAGATATATTAGAAAATGAAAGATGAAACAGAAATAGTCCGCCAAAAAAGGTTAAAGAGG			3872
DB	7261	AGGAA			7202
QY	3873	CCGACCCACCAAAACCTCTTGTTGCGATGACACCAAAAGAAAGCCCAACATGAAAACTTC			3932
DB	7201	GAAAGGAAAGGAG			7142

Oy	3993	TAAAAAGGAGCAAAAAAAAAAATCTGGACCTCCAGCACACAGAGGAGGAGGAAGAAGA	3992
Dd	7141	GAAAGCAAGGAAGAAAAGAAAAGAAAGAAAGAGAAAGAAAAGAAAGAGAAAGA	7082
Oy	3993	AAGCAAAGTGGAATACGGAACAGAACGTCCAAAAGCANAAGCACCGCATGTCAAGGAG	4052
Dd	7081	AAGAAAGCAAGAAATTAGAGAAAGCAAGAACGAAAGAGAGAAAGAGAGAGAAAGAA	7022
Oy	4053	AGCACAGCAGAGA	4065
Dd	7021	AGAAAGACAGAGA	7009

```

RESULT 5
US-09-657-474-7/c
: Sequence 7, Application US/09657474
: Patent No. 6399762
: GENERAL INFORMATION:
: APPLICANT: Chen, H.
: APPLICANT: Prelmer, N.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
: TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
: FILE REFERENCE: 7853-138
: CURRENT APPLICATION NUMBER: US/09/657,474
: CURRENT FILING DATE: 2000-09-07
: PRIOR APPLICATION NUMBER: 09/568,992
: PRIOR FILING DATE: 1999-03-16
: PRIOR APPLICATION NUMBER: 09/236,134
: PRIOR FILING DATE: 1999-01-22
: PRIOR APPLICATION NUMBER: 60/106,056
: PRIOR FILING DATE: 1998-10-28
: PRIOR APPLICATION NUMBER: 60/088,312
: PRIOR FILING DATE: 1998-06-05
: PRIOR APPLICATION NUMBER: 60/078,044
: PRIOR FILING DATE: 1998-03-16
: NUMBER OF SEQ ID NOS: 84
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 7
: LENGTH: 72604
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: modified base
: LOCATION: all n positions
: OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

```

	Query Match	1.2%;	Score 51.4;	DB 4;	Length 72604;
	Best Local Similarity	46.1%;	Pred. No. 0.0076;		
	Matches	172;	Conservative	0;	Mismatches 201; Indels 0; Gaps 0;
OY	3693	GGATGACTTGGACTAAGTTGGTACAGGAACAGAAACCTTAAGGCAGTACGCCAGTCCGAA	3752		
Db	7381	GGATTAACCTTAGGCCAGGGGCTTCTTAGGCTTACCCTGAGCAGCATAGTGGAACCTGGTCCA	7322		
OY	3753	AAGAGCCCATACGGCTTCAGAAATCGATGACACAGCAGTGGCCTAGAGAAAGAGCTCAA	3812		
Db	7321	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	7262		
OY	3813	AGAAGATATATTAGAAAAATGAAGATGAACAGAAATAGTCCGCCAAAAAAGGTTAAAGAGG	3872		
Db	7261	AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA	7202		
OY	3873	CCGACCACCAAAACCTCTTGSTGGAGTATACCCAAAAGAAAGGCCAACAAATGAAACATCTC	3932		
Db	7201	GAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG	7142		
OY	3933	TAAAAAAGGAAGCAAAAAAATATCTGAGCTCCAGCACAGAGAGAGAGGAAGGAAGGAAGGA	3992		
Db	7141	GAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG	7082		
OY	3993	AAGCAAAAGTGGAAATATCGCAACGAAGTCCAAAAAGCAAAACGACGAGTGTTCAAGAGG	4052		

Qy	3724	AAACCTTAAGCGACGTGACGAGACGCGAAGGCGCATACGGCTTCAGAAATCTGATGAA	3788
Db	3305	AAAATTAAACATTTATACACCAAAATPAGGGGAGAAAATTTTGCAGAAAGGAATGATTTCT	3366
Qy	3784	CAGCAGTGGCTGGAGAAAGAGCGGCTCAAGAGATATATTAGAAAATAGAGATGACAG	3844
Db	3305	ACAAAGAAATPACCGATGAATTCGAAAAATGGATGAGTGATPAGGAAACGTGAAAGAAACCTTCA	3422
Qy	3844	AATGATCCGCCAAAAAAGGTAAAAAGAGCGCGACACCAAAACCTCTGGTGGAGGTACA	3902
Db	3425	AAGAAATGCTGAGATATATAGAAAAATTTGAAAGACGCAAAAAGCGCTGCGGCAAAAAAAA	3484
Qy	3904	CCAAAAGAGAGCCCAACATGAAAACTCTAAAAAAGGAGCAAAAATAAAATCTGAGCT	3962
Db	3485	GCAGACAGCGGAGAAAGAAAAGGCTGAGACCTGTAAAAAAGCAGAGAGAGAAAAGAAACGA	3544

RESULT 8
US-09-244-796-17
: Sequence 17, Application US/09244796
: Patent No. 6281344
: GENERAL INFORMATION:
: APPLICANT: Szostrak, Jack W.
: APPLICANT: Szostrak, Richard W.


```

1 TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
2 TITLE OF INVENTION: AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
3 TITLE OF INVENTION: DELIVERING POLYNUCLEOTIDE VACCINES
4 NUMBER OF SEQUENCES: 2
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: Naval Medical Res. & Dev. Cmd.
7 STREET: Bldg. 1, T-12 8901 Wisconsin Ave.
8 CITY: Bethesda
9 STATE: Maryland
10 COUNTRY: USA
11 ZIP: 20889-5606
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.25
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/155,888
19 FILING DATE:
20 CLASSIFICATION: 514
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Spevack, A. David
23 REGISTRATION NUMBER: 24,743
24 REFERENCE/DOCKET NUMBER: N.C. 75,851
25 TELECOMMUNICATION INFORMATION: N.C.
26 TELEPHONE: (202) 295-6759
27 TELEFAX: (202) 295-1022
28 INFORMATION FOR SEQ ID NO: 1:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 5552 base pairs
31 TYPE: nucleic acid
32 STRANDEDNESS: double
33 TOPOLOGY: circular
34 HYPOTHETICAL: NO
35 ANTI-SENSE: NO
36 FEATURE:
37 NAME/KEY: promoter
38 LOCATION: 1..755
39 IDENTIFICATION METHOD: experimental
40 OTHER INFORMATION: /function= "promoter"
41 OTHER INFORMATION: /evidence= EXPERIMENTAL
42 OTHER INFORMATION: /label= CMV-IE
43 OTHER INFORMATION: /note= "This feature acts as a promoter for any
44 OTHER INFORMATION: downstream DNA sequence."
45 OTHER INFORMATION: /citation= ([1])
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 933..2367
49 IDENTIFICATION METHOD: experimental
50 OTHER INFORMATION: /codon_start= 933
51 OTHER INFORMATION: /function= "protein protective against malaria"
52 OTHER INFORMATION: /product= "protein"
53 OTHER INFORMATION: /evidence= EXPERIMENTAL
54 OTHER INFORMATION: /number= 1
55 OTHER INFORMATION: /label= IL2-CSP
56 OTHER INFORMATION: /citation= ([1])
57 PUBLICATION INFORMATION:
58 AUTHORS: Sedegah, Martha
59 AUTHORS: Hedstrom, Richard C.
60 AUTHORS: Hoffman, Stephen L.
61 TITLE: Vaccination with Plasmodium yoelii CS protein
62 TITLE: plasmid DNA protects against malaria
63 JOURNAL: Science
64 PUBLICATION INFORMATION:
65 AUTHORS: Cullen, Bryan R.
66 TITLE: TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
67 TITLE: VIRUS OCCURS VIA A BIMODAL MECHANISM
68 JOURNAL: CELL
69 VOLUME: 46
70 PAGES: 973-982
71 DATE: 26 SEP-1986
72 RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 4732
73 US-08-155-888-1

```

	Query Match	1.0%;	Score 43.6;	DB 3;	Length 5552;
	Best Local Similarity	49.1%;	Pred. No. 0.23;		
	Matches	115;	Conservative	0;	Mismatches 119; Indels 0; Gaps 0;
OY	3771 AGAATCTGATCAACAGCAGTGGCCGTGAAGAAGAGCGCTCAAAGAGATATTATTAGAAAA	3830			
Db	1415 AAAACCCAGAAACAAAAAAAAAGATGATCCCCAAAAGATGGCCAAACAAGATGATCTTCACAAA	1474			
OY	3831 TGAAGATGAACAGCAATAGTCCGCCCAAAAAAGGTTAAAGAGCGGCAGCACCAAAACCCT	3890			
Db	1475 AGAAGAAAAAANAATGATCTTCCAAAAAGAAAGAAAAANAATGATCCCCAAAAAAGATCC	1534			
OY	3891 TGTGTGAGGTRACACCAAAAGAGGCCAACATGAAACTTCTAAAAAAGAGACAAAAA	3950			
Db	1535 TAAAAAAGATGATCCACCCAAAAGAGGCTCAAAAATAATGTAATCAACCGTAGTGCGCAGA	1594			
OY	3951 AAATCTTGAGACCTCCAGCAGCAGAGAGAGAGAAAGAAAGAAAGCAAAATGG	4004			
Db	1595 TGAAATATGTAGATCAAGGCCAGAGACCAACCAAGGCGCAGAGACCAACCAAG	1648			

```

RESULT 12
US-09-434-408-3
: Sequence 3, Application US/09434408
: Patent No. 6440697
: GENERAL INFORMATION:
: APPLICANT: Venezia, Domenick
: APPLICANT: Grossmann, Angelika
: TITLE OF INVENTION: RING FINGER PROTEIN ZAPOB3
: FILE REFERENCE: 98-41
: CURRENT APPLICATION NUMBER: US/09/434,408
: CURRENT FILING DATE: 1999-11-04
: EARLIER APPLICATION NUMBER: US 60/108,258
: EARLIER FILING DATE: 1998-11-12
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 2169
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate polynucleotide sequence of zapob3
: NAME/KEY: misc_feature
: LOCATION: (1)...(2169)
: OTHER INFORMATION: n = A,T,C or G
US-09-434-408-3

```

	Query Match	1.0%; Score 42.4; DB 4; Length 2169;
	Best Local Similarity	23.6%; Pred. No.0.23; Mismatches 173; Indels 0; Gaps 0;
	Matches	76; Conservative 73; Mismatches 173; Indels 0; Gaps 0;
QY	3779 ATGAACAGCAGTGGCTGTGAGGAAAAAGAGGCTCAAGAAGATATATTAGAAATGAAGATG	38388
Db	842 AYACNCARYYTYTTCARCRCRWSNMNSNCCARAARGATGATATHTYNCRRCNCTNARG	901
QY	3839 AACCAATAGTCCGCCCAAAAAGCGTAAAGAGGCCGCACCACCAAACCTTGCTGTGAG	38988
Db	902 ARGCARCRWSMNGYNTNGARCRCRGANTTNMSGARCAYCARMGNCAVYTNGAYCNGCARM	961
QY	3899 GTACACCAAAAAGAGAGCCCAACTTTGAAACTTTTAAAAAAGGAAGCAAAAAAATCTG	39588
Db	962 GNCARMKNYNTCCAGARGCARTYTNARCRACNCGARCARAATAATHTMSNMNSNGNTHCARA	10211
QY	3959 GACCTCCAGCCACGAGAGGAGGAGAAGACAAGAACAAATGTGAATTCGGAACGA	40188
Db	1022 ARITYNYNRCARGATYAATCARBMGNCARAKARAARWBSMWSGARATHTYTNARKSNYNTNGARA	10811
QY	4019 AGTCCAAAAGCAAAACAGCACCCGACTGTCCAAGAGAGCAGACAGAGACAGATCTCCTG	40788
Db	1082 AYGARMGNATHMGNAATGGARGCARTYTNATGSMNATHACICCARGAGARGCNGARBSNYTM	11411
QY	4079 AATCTAGTGCATTTGAATCCAC	4100

FT	exon	378..464
FT		/number= e
FT	exon	/number= 4
FT		465..362
FT		/tag= f
FT	exon	/number= 5
FT		563..669
FT		/tag= e
FT	exon	/number= 6
FT		690..854
FT		/tag= f
FT	exon	/number= 7
FT		855..911
FT		/tag= g
FT	exon	/number= 8
FT		912..1027
FT		/tag= h
FT	exon	/number= 9
FT		1028..1122
FT		/tag= i
FT	exon	/number= 10
FT		1123..1268
FT		/tag= j
FT	exon	/number= 11
FT		1269..1420
FT		/tag= k
FT	exon	/number= 12
FT		1421..1534
FT		/tag= l
FT	exon	/number= 13
FT		1535..1616
FT		/tag= m
FT	exon	/number= 14
FT		1617..1665
FT		/tag= n
FT	exon	/number= 15
FT		1666..1805
FT		/tag= o
FT	exon	/number= 16
FT		1806..1921
FT		/tag= p
FT	exon	/number= 17
FT		1922..2027
FT		/tag= q
FT	exon	/number= 18
FT		2028..2188
FT		/tag= r
FT	exon	/number= 19
FT		2189..2312
FT		/tag= s
FT	exon	/number= 20
FT		2313..2471
FT		/tag= t
FT	exon	/number= 21
FT		2472..2540
FT		/tag= u
FT	exon	/number= 22
FT		2541..2677
FT		/tag= v
FT	exon	/number= 23
FT		2678..2801
FT		/tag= w
FT	exon	/number= 24
FT		2802..3006
FT		/tag= x
FT	exon	/number= 25
FT		3007..3121
FT		/tag= y
FT	exon	/number= 26
FT		3122..3254
FT		/tag= z
FT	exon	/number= 27
FT		3255..3374

Query Match	Best Local Similarity	100.0%	Score 4173;	DB 21;	Length 5271;	Mismatches 0;	Indels 0;	Gaps 0;
Matches 4173;	Conservative	0;						

1 ATGCTCATTCNAAGACTAGACCAATGATGGAATAATTACATATCCCGCTGGCTCAAG 60
66 ATGGCTCATTCNAAGACTAGACCAATGATGGAATAATTACATATCCCGCTGGCTCAAG 125
61 GAATATGAGATAAATATCTAAGAGAGATGTGAGACGATTAAGATGGTGTGAAA 120
126 GAATATGAGATAAATATCTAAGAGAGATGTGAGACGATTAAGATGGTGTGAAA 185
121 ACTTTATGATATGAGACGAGACTCTGAAGAGAAAAGAGCTTATTTAAAGCTACT 180
186 ACTTTATGATATGAGACGAGACTCTGAAGAGAAAAGAGCTTATTTAAAGCTACT 245
181 TTACATCTTGGCTCAGATTTTTTCTCAAGCATCTGCTGAAGATGTTGCTTACTGTA 240
246 TTACATCTTGGCTCAGATTTTTTCTCAAGCATCTGCTGAAGATGTTGCTTACTGTA 305
241 GCCCTGCTCTTGGCTCAGATTTTTTCTCAAGCATCTGCTGAAGATGTTGCTTACTGTA 300
306 GCCCTGCTCTTGGCTCAGATTTTTTCTCAAGCATCTGCTGAAGATGTTGCTTACTGTA 365
301 GATAACTAAAGATATATTTATGTTTATACCAACAGAGTTGAAGGGCTGAGATACA 360
366 GATAACTAAAGATATATTTATGTTTATACCAACAGAGTTGAAGGGCTGAGATACA 425
361 AAGAGCCCAATTCATAGATATTTTATTTACTTGAAGACATTCCTTGGCTCAAGTCA 420
426 AAGAGCCCAATTCATAGATATTTTATTTACTTGAAGACATTCCTTGGCTCAAGTCA 485
421 TATACATATGCTTGAAGTATAGACATAGCAATGAATTTTCCACCAGCTATACGAAC 480
486 TATACATATGCTTGAAGTATAGACATAGCAATGAATTTTCCACCAGCTATACGAAC 545
481 TTATTTAGTTTATTAACAAATGAGCCCAATCAATCAAGATGATGATGATACCTT 540
546 TTATTTAGTTTATTAACAAATGAGCCCAATCAATCAAGATGATGATGATACCTT 605
541 ATGAGCTCTATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTGGATACGTTTGA 600
606 ATGAGCTCTATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTGGATACGTTTGA 665
601 GTAAATCTGTACCTGCTCATTAAGATTTTAAACAAGCAATATGATTTGGCAAAAGCT 660
666 GTAAATCTGTACCTGCTCATTAAGATTTTAAACAAGCAATATGATTTGGCAAAAGCT 725
661 TTACAGAGAGAGACTCAAGCTATGAGCATATATACACTTTTAAATAGAT 720
726 TTACAGAGAGAGACTCAAGCTATGAGCATATATATACACTTTTAAATAGAT 785
721 CTGATGCTTGGGAAAACATCTATCAGCATTTGTCAGAGCATGCTTTGACTTAATTTG 780
786 CTGATGCTTGGGAAAACATCTATCAGCATTTGTCAGAGCATGCTTTGACTTAATTTG 845
781 GAGCTCTACAAATATGATATGATTTGCTGCTCTGTTTTAACCCAGCTGGAATTTAA 840
846 GAGCTCTACAAATATGATATGATTTGCTGCTCTGTTTTAACCCAGCTGGAATTTAA 905
841 TTAAGAGACAAATGATAGAGAGGCTTACAAAGTGTGTTAACTATGCAAAATTTGTT 900
906 TTAAGAGACAAATGATAGAGAGGCTTACAAAGTGTGTTAACTATGCAAAATTTGTT 965
901 GGGGCAAGAGATTCAGATTTGGCTTCTCAAAAACAGCATTGGCAGTGTGCTTGGGC 960
966 GGGGCAAGAGATTCAGATTTGGCTTCTCAAAAACAGCATTGGCAGTGTGCTTGGGC 1025
961 AGTTTTAATGATATCCATGTACCAATCCGCTGGAATGTGAAAATTTGCTAGCCATTT 1020
1026 AGTTTTAATGATATCCATGTACCAATCCGCTGGAATGTGAAAATTTGCTAGCCATTT 1085
1021 CTCATGAACATCCGATTTAGCAAAAGATCTTAACAGATATCTTAAGGTAGGTCACAT 1080
1086 CTCATGAACATCCGATTTAGCAAAAGATCTTAACAGATATCTTAAGGTAGGTCACAT 1145
1081 GACCCTGAGGAAGCTATTAGACATGATGTTATGTTGTCATATAGTTACAGCTGCTAAAAG 1140

1146 GACCCTGAGGAAGCTATTAGACATGATGTTATGTTGTCATATAGTTACAGCTGCTAAAAG 1205
1141 GATATTTCTTGTGTCATATGATCATTACTTAATTTTGTGAGAGAGCAACATTTAGACAAA 1200
1206 GATATTTCTTGTGTCATATGATCATTACTTAATTTTGTGAGAGAGCAACATTTAGACAAA 1265
1201 CGATGAGAGATACGAAAGAACCCATGATGAGGACTTGGCCAAATTTATAGAAAATATGCT 1260
1266 CGATGAGAGATACGAAAGAACCCATGATGAGGACTTGGCCAAATTTATAGAAAATATGCT 1325
1261 TTACAGTACAGAGCTGGAAGAAAGATGCTCAAAACAGATAGATGATCAAGACAAATTTG 1320
1326 TTACAGTACAGAGCTGGAAGAAAGATGCTCAAAACAGATAGATGATCAAGACAAATTTG 1385
1321 CTACATATATATTTATCAAAATATGATATGATGATGATGATGATGATGATGATGATGAT 1380
1386 CTACATATATATTTATCAAAATATGATATGATGATGATGATGATGATGATGATGATGAT 1445
1381 CAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
1446 CAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1505
1441 TATGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
1506 TATGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1565
1501 CTGCTCGCATCAAGATTAAGATTTGCTTGAATTTGATTAAGCAACCAAAACAGATGCC 1560
1566 CTGCTCGCATCAAGATTAAGATTTGCTTGAATTTGATTAAGCAACCAAAACAGATGCC 1625
1561 AGTGTCAAGGGCATTTTCAAAAGTGAATGATTAAGCAAAATTTTCTGATCCTGCT 1620
1626 AGTGTCAAGGGCATTTTCAAAAGTGAATGATTAAGCAAAATTTTCTGATCCTGCT 1685
1621 AAGGCTCAGATTTTCAAGATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1680
1686 AAGGCTCAGATTTTCAAGATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1745
1681 AAGCATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1740
1746 AAGCATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1805
1741 CGTGAATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1800
1806 CGTGAATTAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGAT 1865
1801 ATCAAGTTTCTTGTGAGAGATAGACACTGCTGACATAGATATCCGAATCTATCAGTCT 1860
1866 ATCAAGTTTCTTGTGAGAGATAGACACTGCTGACATAGATATCCGAATCTATCAGTCT 1925
1861 CTTATTTAAACAAGTAAACAAATCAATATGATGATGATGATGATGATGATGATGATGATGAT 1920
1926 CTTATTTAAACAAGTAAACAAATCAATATGATGATGATGATGATGATGATGATGATGATGAT 1985
1921 CCAACTGATCAAGCATGAGAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1980
1986 CCAACTGATCAAGCATGAGAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2045
1981 CCAACTGATCAAGCATGAGAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2040
2046 CCAACTGATCAAGCATGAGAGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2105
2041 GATGAAAAAGTAAAGAGAGGCTGCTACAAATTTTCAAAAACAGAGAAAGCAAAATTTGAA 2100
2106 GATGAAAAAGTAAAGAGAGGCTGCTACAAATTTTCAAAAACAGAGAAAGCAAAATTTGAA 2165
2101 GAGGATTTTCCACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
2166 GAGGATTTTCCACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2225
2161 GGACCCCCCGTCAAGCCAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220

Db	2226	GGACCCCCCGCTCAGCCCAATATGCCATTCTGATTCATGCGGATATTTTCTAGTAA	2285
QY	2221	GAGACCCAGTTTGGACAGATATTTGAGCCTCTGCATPAAGACCTAGATCCAGCAACCTG	2280
Db	2286	GAGACCCAGTTTGGACAGATATTTGAGCCTCTGCATPAAGACCTAGATCCAGCAACCTG	2345
QY	2281	GAAATCTCATPAACCACTTGGTTACTATATTTGGTCTATATTTGCTTCCTTGCACCTGATCAA	2340
Db	2346	GAACATCTCATPAACCACTTGGTTACTATATTTGGTCTATATTTGCTTCCTTGCACCTGATCAA	2405
QY	2341	TTTGCTGCTCCTTGGAAATCTTGGGTAAGCTCTTCAATGTGAAGAATCTCTATGAAT	2400
Db	2406	TTTGCTGCTCCTTGGAAATCTTGGGTAAGCTCTTCAATGTGAAGAATCTCTATGAAT	2465
QY	2401	GATCGGCTCCAGGGAAAAAGACAACCTAACTTTGGGTTCCAGATGAAGAATCTCTCT	2460
Db	2466	GATCGGCTCCAGGGAAAAAGACAACCTAACTTTGGGTTCCAGATGAAGAATCTCTCTCT	2525
QY	2461	GAGCAATNGTCATAAATTTCAAGCTATTTAAATGATGGTTCGATGGCTACTTGGAAATGAA	2520
Db	2526	GAGCAATNGTCATAAATTTCAAGCTATTTAAATGATGGTTCGATGGCTACTTGGAAATGAA	2585
QY	2521	AATATATCACAATAATCAGAACTTCACTTAAGTTTGGTAACAAATAATTTGATCATAGT	2580
Db	2586	AATATATCACAATAATCAGAACTTCACTTAAGTTTGGTAACAAATAATTTGATCATAGT	2645
QY	2581	GATGAGACTTGTACAGAAACAGGGGAAAAATTAGTAACCAATATGTACGCTTGAGACTT	2640
Db	2646	GATGAGACTTGTACAGAAACAGGGGAAAAATTAGTAACCAATATGTACGCTTGAGACTT	2705
QY	2641	GCTGCTGGAGATGCTATTTGTGAACTGCGACAAACCCCTGTACCATGGAATCATACACA	2700
Db	2706	GCTGCTGGAGATGCTATTTGTGAACTGCGACAAACCCCTGTATCCATGGAATCATACACA	2765
QY	2701	TTAGAACAAATTCAGCTATGTGCTATTAGCTATCAACGATGATCTATCAAGTAAGACAA	2760
Db	2766	TTAGAACAAATTCAGCTATGTGCTATTAGCTATCAACGATGATCTATCAAGTAAGACAA	2825
QY	2761	GTGTTTCCCAAGAACTTCACAAAGGCTTTCCCGTTTACGGCTTCCACTTGAATATATG	2820
Db	2826	GTGTTTCCCAAGAACTTCACAAAGGCTTTCCCGTTTACGGCTTCCACTTGAATATATG	2885
QY	2821	GCAATCTGTGCTTTGTGCAAAAGATCCGTAAAGGAGAAACAGTCATGCTAAGGCCAA	2880
Db	2886	GCAATCTGTGCTTTGTGCAAAAGATCTTCTAAAGGAGAAACAGTCATGCTAAGGCCAA	2945
QY	2881	TGTTTGGTGAAAAATATPAAATGTAAAGGGGAGATCTGAAGCAGACATGCAGCTGTATGT	2940
Db	2946	TGTTTGGTGAAAAATATPAAATGTAAAGGGGAGATCTGAAGCAGACATGCAGCTGTATGT	3005
QY	2941	GAATAATTTATTTGCTCTTCTACACAGAGTATGTGTTCCATATPACAAATTCACCTTTGGCA	3000
Db	3006	GAATAATTTATTTGCTCTTCTACACAGAGTATGTGTTCCATATPACAAATTCACCTTTGGCA	3065
QY	3001	CATGACCCAGATTATGTCAAGTACAGATATTTGAACAACCTTAAAGATGTTAAAGATGT	3060
Db	3066	CATGACCCAGATTATGTCAAGTACAGATATTTGAACAACCTTAAAGATGTTAAAGATGT	3125
QY	3061	CTTTGGTTTGTCTGCGAATATPAAATGAGCTPAAAAATGAATAACAGTCACGCTTTTATTC	3120
Db	3126	CTTTGGTTTGTCTGCGAATATPAAATGAGCTPAAAAATGAATAACAGTCACGCTTTTATTC	3185
QY	3121	AGAAAGATGTAGAAAAATTTTAAACAAACAAAGATGCCCAAGGACCAAGATGATGCAAA	3180
Db	3186	AGAAAGATGTAGAAAAATTTTAAACAAACAAAGATGCCCAAGGACCAAGATGATGCAAA	3245
QY	3181	ATGAATGAAAAACGTGTACAGTGTGTGTATGTTGCCATGAATATCATCATGTCAAAAGGT	3240
Db	3246	ATGAATGAAAAACGTGTACAGTGTGTGTATGTTGCCATGAATATCATCATGTCAAAAGGT	3305
QY	3241	ACTACATACAGTTTGGATCTCTPAAACCCCGGACTACACAGCTCGTTTCTTCACTCAA	3300
Db	3306	ACTACATACAGTTTGGATCTCTPAAACCCCGGACTACACAGCTCGTTTCTTCACTCAA	3365

QY	3301	CCTGCAAGAAATTTTCAGTAACACCAAAATATCTGCCTCCTGAAATGAATCATTTTTC	3360
Db	3366	CCTGCAAGAAATTTTCAGTAACACCAAAATATCTGCCTCCTGAAATGAATCATTTTTC	3425
QY	3361	ACTCCTGGAAAACCTTAACAAACCAATGCTCTAGAGCTGTTAACAAACCCACTTTCATCA	3420
Db	3426	ACTCCTGGAAAACCTTAACAAACCAATGCTCTAGAGCTGTTAACAAACCCACTTTCATCA	3485
QY	3421	GCAGGCAAGCAATCTCAGACCAAAATCATCACGAATGGAACCTGTAAAGCAATGCACAGC	3480
Db	3486	GCAGGCAAGCAATCTCAGACCAAAATCATCACGAATGGAACCTGTAAAGCAATGCACAGC	3545
QY	3481	AGCTCAATTCAGAGCTCTCTTGGAAGATTAAGGGAGGCTTGATAGTTCTGAATGAT	3540
Db	3546	AGCTCAATTCAGAGCTCTCTTGGAAGATTAAGGGAGGCTTGATAGTTCTGAATGAT	3605
QY	3541	CACAGTGAATTAAGAAATTTACACAATGCTTCACCTTGGCGGGGAAAAAAGTACAG	3600
Db	3606	CACAGTGAATTAAGAAATTTACACAATGCTTCACCTTGGCGGGGAAAAAAGTACAG	3665
QY	3601	AGAGAGCACTGTGATCTTTAAGGTCTGAATTTGGAAGCCTAGAGCAGAGAAAAAACG	3660
Db	3666	AGAGAGCACTGTGATCTTTAAGGTCTGAATTTGGAAGCCTAGAGCAGAGAAAAAACG	3725
QY	3661	CCCGTCACAGAACAGAGAGAGAAATTAAGTATGATGACTTGACTTAAGTTGCTACAGAA	3720
Db	3726	CCCGTCACAGAACAGAGAGAGAAATTAAGTATGATGACTTGACTTAAGTTGCTACAGAA	3785
QY	3721	CAGAAACCTTAAAGGAGTACAGGAAGTGGGAAAGAGGCGATACGGCTTCAGAAATCTGAT	3780
Db	3786	CAGAAACCTTAAAGGAGTACAGGAAGTGGGAAAGAGGCGATACGGCTTCAGAAATCTGAT	3845
QY	3781	GAACAGCAGTGGCTTGAGAAAAAGAGGCTCAAAAGAAATATATTAGAAAATGAAGATGAA	3840
Db	3846	GAACAGCAGTGGCTTGAGAAAAAGAGGCTCAAAAGAAATATATTAGAAAATGAAGATGAA	3905
QY	3841	CAGATTAATCCGCCCAAAAAAGGGTTAAAGAGGCGCACCAACCAAACTCTTGTTGGAGGT	3900
Db	3906	CAGATTAATCCGCCCAAAAAAGGGTTAAAGAGGCGCACCAACCAAACTCTTGTTGGAGGT	3965
QY	3901	ACACCAAAAGAAAGAGCCCAACAATGAAAACCTTCTAATAAAAGGAAGCAAAAAAATCTGGA	3960
Db	3966	ACACCAAAAGAAAGAGCCCAACAATGAAAACCTTCTAATAAAAGGAAGCAAAAAAATCTGGA	4025
QY	3961	CCTCCAGCACAGAGAGAGAGAGAAAGAAAGCAAAAGTGGAAATACGGAAACAGAG	4020
Db	4026	CCTCCAGCACAGAGAGAGAGAGAAAGAAAGCAAAAGTGGAAATACGGAAACAGAG	4085
QY	4021	TCCAAAAGCAAAACAGCACCGAGTGTCAAGGAGAGCAGCAGCAGAGCGAATCTCTGAA	4080
Db	4086	TCCAAAAGCAAAACAGCACCGAGTGTCAAGGAGAGCAGCAGCAGAGCGAATCTCTGAA	4145
QY	4081	TCTAGTGAATTAAGTTCACACAGTCCACACCACAGAAAGGAGCGAGAGACATCAAAA	4140
Db	4146	TCTAGTGAATTAAGTTCACACAGTCCACACCACAGAAAGGAGCGAGAGACATCAAAA	4205
QY	4141	ACGCCATCACATTCACAACCAAAAAAATGTG	4173
Db	4206	ACGCCATCACATTCACAACCAAAAAAATGTG	4238
RESULT 2			
AAA28052			
ID	AAA28052 standard; cDNA; 5355 BP.		
XX	AAA28052;		
AC	XX		
DT	01-DEC-2000 (first entry)		
XX	XX		
DE	Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.		
XX	XX		
Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;			

KW chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
 KW diagnosis; treatment; cytostatic; human; ss.
 OS Homo sapiens.
 XX
 XX Key
 FT exon
 FT Location/Qualifiers
 FT 85..130
 FT /tag= a
 FT /number= 1
 FT 131..257
 FT /tag= b
 FT /number= 2
 FT 150..4325
 FT /tag= c
 FT /product= "AS3"
 FT /note= "Androgen shutoff gene 3 protein"
 FT 258..461
 FT /tag= d
 FT /number= 3
 FT 462..548
 FT /tag= e
 FT /number= 4
 FT 549..646
 FT /tag= f
 FT /number= 5
 FT 647..773
 FT /tag= e
 FT /number= 6
 FT 774..938
 FT /tag= f
 FT /number= 7
 FT 939..995
 FT /tag= g
 FT /number= 8
 FT 996..1111
 FT /tag= h
 FT /number= 9
 FT 1112..1206
 FT /tag= i
 FT /number= 10
 FT 1207..1352
 FT /tag= j
 FT /number= 11
 FT 1353..1504
 FT /tag= k
 FT /number= 12
 FT 1505..1618
 FT /tag= l
 FT /number= 13
 FT 1619..1700
 FT /tag= m
 FT /number= 14
 FT 1701..1749
 FT /tag= n
 FT /number= 15
 FT 1750..1889
 FT /tag= o
 FT /number= 16
 FT 1890..2005
 FT /tag= p
 FT /number= 17
 FT 2006..2111
 FT /tag= q
 FT /number= 18
 FT 2112..2272
 FT /tag= r
 FT /number= 19
 FT 2273..2396
 FT /tag= s
 FT /number= 20
 FT 2397..2555
 FT /tag= t
 FT /number= 21

FT exon
 FT 2556..2624
 FT /tag= u
 FT /number= 22
 FT 2625..2761
 FT /tag= v
 FT /number= 23
 FT 2672..2885
 FT /tag= w
 FT /number= 24
 FT 2886..3090
 FT /tag= x
 FT /number= 25
 FT 3091..3205
 FT /tag= y
 FT /number= 26
 FT 3206..3338
 FT /tag= z
 FT /number= 27
 FT 3339..3458
 FT /tag= aa
 FT /number= 28
 FT 3459..3521
 FT /tag= ab
 FT /number= 29
 FT 3522..3667
 FT /tag= ac
 FT /number= 30
 FT 3668..3773
 FT /tag= ad
 FT /number= 31
 FT 3774..4213
 FT /tag= ae
 FT /number= 32
 FT 4214..4438
 FT /tag= af
 FT /number= 33
 FT 4314..4408
 FT /tag= ag
 FT /note= "Destabilizing signal"
 FT 4439..5337
 FT /tag= ah
 FT /number= 34
 FT 5312..5317
 FT /tag= ai
 FT 5333..5337
 FT /tag= aj
 FT /note= "Poly adenylation cleavage site"
 FT
 PN exon
 PD
 XX 31-AUG-2000.
 XX 24-FEB-2000; 2000WO-US04732.
 XX PF
 XX 24-FEB-1999; 99US-0121461.
 XX PR
 XX (TUFT) TUFTS COLLEGE.
 PA
 XX Soto AM, Sonnenschein C, Geck P, Szelei J;
 XX PI WPL; 2000-565451/52.
 XX DR P-PSDB; AAY94702.
 XX DR
 XX
 XX New human androgen-induced tumor suppressor cDNA sequence termed
 PT 'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
 PT diagnosis and treatment of prostate cancer -
 XX
 XX Example 4; Fig 6; 152pp; English.
 PS
 CC This invention relates to a human androgen-induced tumour suppressor
 CC cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
 CC located on chromosome 13 at position 13q12-13q. AS3 has a role in
 CC inhibiting cell proliferation and use as a marker for the efficient

CC diagnosis and treatment of prostate cancer. The invention includes AS3
CC cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC cell transfected with the expression vector, and a method for producing
CC an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC cytoskeletal activity, and acts to suppress cell proliferation. The AS3
CC gene is useful as a marker for the efficient diagnosis and treatment of
CC prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC antisense agents for sequence specific modulation of gene expression. The
CC AS3 protein may be used in the treatment of disorders caused by aberrant
CC modification or mutation of a gene encoding an AS3 protein, misregulation
CC of the AS3 gene or aberrant post-translational modification of the AS3
CC protein. This sequence represents the human AS3 cDNA sequence with an
CC additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC compared with the claimed AS3 cDNA sequence AAA28051.

XX Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Query Match 100.0%; Score 4173; DB 21; Length 5355;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCATTAAGAGCTAGAGCAATGATGGAATAATTCATATCCGCTGGGCTCAAG 60
DB 150 ATGGCTCATTAAGAGCTAGAGCAATGATGGAATAATTCATATCCGCTGGGCTCAAG 209
QY 61 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGAGACAGATTAAAGATGTTGTGAAA 120
DB 210 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGAGACAGATTAAAGATGTTGTGAAA 269
QY 121 ACTTTTATGATATGAGACAGAGCTCTGGAAGAAAAGAGCTTTATTTAACTTACT 180
DB 270 ACTTTTATGATATGAGACAGAGCTCTGGAAGAAAAGAGCTTTATTTAACTTACT 329
QY 181 TTACATCTTGGTGTAGATTTTCTCAGACATCTGTTAAGATGTTGCTTACTGTGTA 240
DB 330 TTACATCTTGGTGTAGATTTTCTCAGACATCTGTTAAGATGTTGCTTACTGTGTA 389
QY 241 GCGTGTGCTTGGTGTAGATTTTCTCAGACATCTGTTAAGATGTTGCTTACTGTGTA 300
DB 390 GCGTGTGCTTGGTGTAGATTTTCTCAGACATCTGTTAAGATGTTGCTTACTGTGTA 449
QY 301 GATAAATCAAGAGATATTTATGTTTATACAGACAGTGTGAAGGGCTGAGAGATACA 360
DB 450 GATAAATCAAGAGATATTTATGTTTATACAGACAGTGTGAAGGGCTGAGAGATACA 509
QY 361 AAGAGCCCAATTCATAGTATTTTATTTACTTGAACATTTGCTTGGTCAAGTCA 420
DB 510 AAGAGCCCAATTCATAGTATTTTATTTACTTGAACATTTGCTTGGTCAAGTCA 569
QY 421 TATAACATATGCTTTGAGTATGAGAGATGAGCAATGAAATTTTCCACGAGCTATACAGAAC 480
DB 570 TATAACATATGCTTTGAGTATGAGAGATGAGCAATGAAATTTTCCACGAGCTATACAGAAC 629
QY 481 TTATTTTCAAGTATTAACAAATGCGCACATACAGAAAGTCCATATGACATAGTAGACCTT 540
DB 630 TTATTTTCAAGTATTAACAAATGCGCACATACAGAAAGTCCATATGACATAGTAGACCTT 689
QY 541 ATGAGCTCTATTTATTTGAGAGTGTATPACAGTGTCTCAGAGCTTTTGGATACGGTTTAA 600
DB 690 ATGAGCTCTATTTATTTGAGAGTGTATPACAGTGTCTCAGAGCTTTTGGATACGGTTTAA 749
QY 601 GTAATCTGTACCTGCTCATPAAGAAATTTAAACAAGCAATATGATTTGGCAAAAGCT 660
DB 750 GTAATCTGTACCTGCTCATPAAGAAATTTAAACAAGCAATATGATTTGGCAAAAGCT 809
QY 661 TTACTGAAGAGAGACAGCTCAAGCTATATGAGCATATATACACATTTTAAATACAGTGT 720
DB 810 TTACTGAAGAGAGACAGCTCAAGCTATATGAGCATATATATACACATTTTAAATACAGTGT 869
QY 721 CTGATGCTTTGGGAAAACATCTATCAGAGATTTGTCAGAGCAATGCTTTGACCTTAATTTTG 780
DB 870 CTGATGCTTTGGGAAAACATCTATCAGAGATTTGTCAGAGCAATGCTTTGACCTTAATTTTG 929

QY 781 GAGCTCTACATATATGATAGTCAATTTGCTGCTCTGTGTTTACCACAGCTTGAATTTAA 840
DB 930 GAGCTCTACATATATGATAGTCAATTTGCTGCTCTGTGTTTACCACAGCTTGAATTTAA 989
QY 841 TTAAAGAGCAATGATATGAGAGAGCGCTTCAAGTGTGTAACTACTGCGAAAATGTTT 900
DB 990 TTAAAGAGCAATGATATGAGAGAGCGCTTCAAGTGTGTAACTACTGCGAAAATGTTT 1049
QY 901 GGGGCAAGAGATTCAGATATGAGTGTCTTCAAAACAAGCAAGCTTTGGCAGTCTTCTGGG 960
DB 1050 GGGGCAAGAGATTCAGATATGAGTGTCTTCAAAACAAGCAAGCTTTGGCAGTCTTCTGGG 1109
QY 961 AGTTTATGATATTCATATGATACCAATCCGCTGGAATGTGTGAATTTGCTTACCATTT 1020
DB 1110 AGTTTATGATATTCATATGATACCAATCCGCTGGAATGTGTGAATTTGCTTACCATTT 1169
QY 1021 CTCATGAAACCATCCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGAGAGTCAAT 1080
DB 1170 CTCATGAAACCATCCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGAGAGTCAAT 1229
QY 1081 GACCCGTAGAGAGCTATTTAGCAATGATGTTATGTTGTCATATGATTTACAGCTGCTAAAAG 1140
DB 1230 GACCCGTAGAGAGCTATTTAGCAATGATGTTATGTTGTCATATGATTTACAGCTGCTAAAAG 1289
QY 1141 GATATTTCTTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAAACATTAGACAAA 1200
DB 1290 GATATTTCTTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAAACATTAGACAAA 1349
QY 1201 CGATGAGAGATGAGCAAAAGAGCCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1260
DB 1350 CGATGAGAGATGAGCAAAAGAGCCATGATGGAGCTTCCCAAAATTTTAAAGAAATATGCT 1409
QY 1261 TTACAGTCAGAGCTGGAAGAAAGATGTCGCAAAACAGATAGCATGATCAAAAGACAAATTTG 1320
DB 1410 TTACAGTCAGAGCTGGAAGAAAGATGTCGCAAAACAGATAGCATGATCAAAAGACAAATTTG 1469
QY 1321 CTACATATATATTTTCAAAATATGATATGATGACTGCTGTTGAAGAGATTTTGGT 1380
DB 1470 CTACATATATATTTTCAAAATATGATATGATGACTGCTGTTGAAGAGATTTTGGT 1529
QY 1381 CAATACATATGTTCTCTCAAAATTTTGAAGATGACAGAGAGATGAAATGCTATATTTACTTGT 1440
DB 1530 CAATACATATGTTCTCTCAAAATTTTGAAGATGACAGAGAGATGAAATGCTATATTTACTTGT 1589
QY 1441 TATGCCACACTGATTTTAAATGCTGTGAAGCAATTTGAATGATGGAATGTCAAAAT 1500
DB 1590 TATGCCACACTGATTTTAAATGCTGTGAAGCAATTTGAATGATGGAATGTCAAAAT 1649
QY 1501 CTGCTCCGACATCAAGTAAAGATTTGCTGTGACTTGTATTAAGCAACCCAAACAGATGCC 1560
DB 1650 CTGCTCCGACATCAAGTAAAGATTTGCTGTGACTTGTATTAAGCAACCCAAACAGATGCC 1709
QY 1561 AGTGCAAGGCCATATTTTCAAAAGTGTGTTTTCAGAAATTTTACCGATCTGGT 1620
DB 1710 AGTGCAAGGCCATATTTTCAAAAGTGTGTTTTCAGAAATTTTACCGATCTGGT 1769
QY 1621 AAGGCTCAGATTTTCATGAAGAAATTCACACAGGCTGTAGAAGATGAGAAAATAGA 1680
DB 1770 AAGGCTCAGATTTTCATGAAGAAATTCACACAGGCTGTAGAAGATGAGAAAATAGA 1829
QY 1681 AAGCAGTTAAGAGTACTTGTATGTCACAATGCTGCTGCAAGCAGGCTGAAGGTTGTGTG 1740
DB 1830 AAGCAGTTAAGAGTACTTGTATGTCACAATGCTGCTGCAAGCAGGCTGAAGGTTGTGTG 1889
QY 1741 CGTGAATTAATTAAGAGTGTGGGACCCCAAGAGCTTACAAATTCCTTCTGGAATG 1800
DB 1890 CGTGAATTAATTAAGAGTGTGGGACCCCAAGAGCTTACAAATTCCTTCTGGAATG 1949
QY 1801 ATCAAGTTTCTCTTGGAGAGATAGCACTGTGACATATAGATCCGATATCATGCTGT 1860
DB 1950 ATCAAGTTTCTCTTGGAGAGATAGCACTGTGACATATAGATCCGATATCATGCTGT 2009
QY 1861 CTATTTAAACAAGTGAACAAATCAATATGATGAGAAACAGCATGATGAGGTTGT 1920

Db 2010 CTTATTAAACAAAGTGAACAAATCAATAGATGGAACAGCATGATGAAATGAGGGTGT 2069
QY 1921 CCAACTGATCAAGCATCAGAGAGGCTTGAACCTGTTAAGGACTCTCATTTACAT 1980
Db 2070 CCAACTGATCAAGCATCAGAGAGGCTTGAACCTGTTAAGGACTCTCATTTACAT 2129
QY 1981 CCCATCTCATTTCTGCTGTAACCAATTTGAATTCATTTAGCTGCTGTAACCAATGAT 2040
Db 2130 CCCATCTCATTTCTGCTGTAACCAATTTGAATTCATTTAGCTGCTGTAACCAATGAT 2189
QY 2041 GATGAAAAAGTAGCAGAGCTGCATTCAAATTTTCAAAAACACAGAGCAAAATTTGAA 2100
Db 2190 GATGAAAAAGTAGCAGAGCTGCATTCAAATTTTCAAAAACACAGAGCAAAATTTGAA 2249
QY 2101 GAGGATTTTCCACACATCAGATCAGCTTGCCTGCTTTTACATTCACAAATCTAAAAAA 2160
Db 2250 GAGGATTTTCCACACATCAGATCAGCTTGCCTGCTTTTACATTCACAAATCTAAAAAA 2309
QY 2161 GGAACCCGCCGTCAGGCAAAATATGCCATTCATTTGATTCATGCGATATTTTCTAGTAA 2220
Db 2310 GGAACCCGCCGTCAGGCAAAATATGCCATTCATTTGATTCATGCGATATTTTCTAGTAA 2369
QY 2221 GAGACCCAGTTTGCACAGATATTTTGAGCTCTGCATTAAGCCTAGATCCAGCAACCTG 2280
Db 2370 GAGACCCAGTTTGCACAGATATTTTGAGCCTCTGCATTAAGCCTAGATCCAGCAACCTG 2429
QY 2281 GAACATCTCATTAACACATGTTGTTACTATTTGATGATATTTGCTGCTTGCACCTGATCAA 2340
Db 2430 GAACATCTCATTAACACATGTTGTTACTATTTGATGATATTTGCTGCTTGCACCTGATCAA 2489
QY 2341 TTTTGCTGCTCTTGAATCTTGGTGAAGTCTTCTTGAAGATCTTCTCATGAT 2400
Db 2490 TTTTGCTGCTCTTGAATCTTGGTGAAGTCTTCTTGAAGATCTTCTCATGAT 2549
QY 2401 GATGCGCTTCCAGGAAAAAGACAATACTTGGTTCGACATGAAGAATATCTCCT 2460
Db 2550 GATGCGCTTCCAGGAAAAAGACAATACTTGGTTCGACATGAAGAATATCTCCT 2609
QY 2461 GAGACATGCTCAAAATTTAGGCTATTTAAATGATGTTGATGCTGCTTGAAGTAA 2520
Db 2610 GAGACATGCTCAAAATTTAGGCTATTTAAATGATGTTGATGCTGCTTGAAGTAA 2669
QY 2521 AATTAATCAGTAAATCAGAACTTCTACCTTAAGTGTCTTAACAACAATATTTGATGAT 2580
Db 2670 AATTAATCAGTAAATCAGAACTTCTACCTTAAGTGTCTTAACAACAATATTTGATGAT 2729
QY 2581 GATGAGACTTGCACAGAGGAAAAATTAGTAAACCAATATGTCACCTGAGACTT 2640
Db 2730 GATGAGACTTGCACAGAGGAAAAATTAGTAAACCAATATGTCACCTGAGACTT 2789
QY 2641 GCTGCTGAGAGTCTATTTGAGAGCTGCGACAAGAACCCGTTTACCATGAAATCATTACA 2700
Db 2790 GCTGCTGAGAGTCTATTTGAGAGCTGCGACAAGAACCCGTTTACCATGAAATCATTACA 2849
QY 2701 TTAAGAACATATCAGCTATGTCATTAAGTATTAAGGATGATGATCAAGTAAAGCAA 2760
Db 2850 TTAAGAACATATCAGCTATGTCATTAAGTATTAAGGATGATGATCAAGTAAAGCAA 2909
QY 2761 GTGTTGGCCAGAAACTTGCACAAAGGCTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
Db 2910 GTGTTGGCCAGAAACTTGCACAAAGGCTTCCCGTTTACGGCTTCCACTTGAGTATATG 2969
QY 2821 GCAATCTGTCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAGAGCTCATGCTAGACAA 2880
Db 2970 GCAATCTGTCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAGAGCTCATGCTAGACAA 3029
QY 2881 TGTGTTGGTGAATAATTAATGTAAGGCGGAGATCTGTAAGCAGCATGCAAGCTGTAGT 2940
Db 3030 TGTGTTGGTGAATAATTAATGTAAGGCGGAGATCTGTAAGCAGCATGCAAGCTGTAGT 3089
QY 2941 GAAAAATTTATGCTCTCTACAGAGTATGTTCCATATACATTCACCTTTGGCA 3000

Db 3090 GAAAAATTTATGCTCTCTTACAGAGTATGTTCTTCATATATCAATTCACCTTTGGCA 3149
QY 3001 CATGACCCAGATTTATGCAAAAGTACAGATATTTGAACAACCTTAAGATGTTAAAGATGT 3060
Db 3150 CATGACCCAGATTTATGCAAAAGTACAGATATTTGAACAACCTTAAGATGTTAAAGATGT 3209
QY 3061 CTTTGGTTTGTCTGCAAAATATTTAATGGCTAAATAATGAAATTAACAGTCACGCTTTATC 3120
Db 3210 CTTTGGTTTGTCTGCAAAATATTTAATGGCTAAATAATGAAATTAACAGTCACGCTTTATC 3269
QY 3121 AGAAAGATGTTAATAAATATTAACAACAAAGATGCCAAGACCAAGTATGCAAAA 3180
Db 3270 AGAAAGATGTTAATAAATATTAACAACAAAGATGCCAAGACCAAGTATGCAAAA 3329
QY 3181 ATGAATGAAAAACCTGTACACTGTGTGATGCTGATGATGATGATGATGATGATGATGAT 3240
Db 3330 ATGAATGAAAAACCTGTACACTGTGTGATGCTGATGATGATGATGATGATGATGATGAT 3389
QY 3241 ACTACATACAGTTTGGAAATCTCTTAAAGACCCGGTACTACAGCTGCTTCTTACTGCA 3300
Db 3390 ACTACATACAGTTTGGAAATCTCTTAAAGACCCGGTACTACAGCTGCTTCTTACTGCA 3449
QY 3301 CCTGACAAAGATTTCAAGTAAACACCAAAATATTTGCTGCTCTGTAATGAAATGATTTTC 3360
Db 3450 CCTGACAAAGATTTCAAGTAAACACCAAAATATTTGCTGCTCTGTAATGAAATGATTTTC 3509
QY 3361 ACTCTGGAAGAACCTTAAACCAACCAATGTTCTAGAGAGCTGTTAAACAGCCACTTTCATCA 3420
Db 3510 ACTCTGGAAGAACCTTAAACCAACCAATGTTCTAGAGAGCTGTTAAACAGCCACTTTCATCA 3569
QY 3421 GCAGGCAAGCAATCTCAGACCAATATCAGATGAAATGAAACCTTAAGCAATGCAAGCAGC 3480
Db 3570 GCAGGCAAGCAATCTCAGACCAATATCAGATGAAATGAAACCTTAAGCAATGCAAGCAGC 3629
QY 3481 AGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGGGAGGCTTATGTTGTTGAAGTATGAT 3540
Db 3630 AGCTCAAAATCCAAGCTCTCTGGAAGAAATTAAGGGAGGCTTATGTTGTTGAAGTATGAT 3689
QY 3541 CACAGTGAATAATGAAGTTTACACAAATGCTTTCACCTTTGCGGGGAAAAAAGTGCACAG 3600
Db 3690 CACAGTGAATAATGAAGTTTACACAAATGCTTTCACCTTTGCGGGGAAAAAAGTGCACAG 3749
QY 3601 AGAGACGACTCTGATCTTGTAGGCTGGAATTTGGAGAGCCTTGAAGCAGGAAAAAAGC 3660
Db 3750 AGAGACGACTCTGATCTTGTAGGCTGGAATTTGGAGAGCCTTGAAGCAGGAAAAAAGC 3809
QY 3661 CCGGTCAAGAAACAGGAGGAGAAATTAAGTATGATGATGATGATGATGATGATGATGAT 3720
Db 3810 CCGGTCAAGAAACAGGAGGAGAAATTAAGTATGATGATGATGATGATGATGATGATGAT 3869
QY 3721 CAGAAACCTTAAAGGAGTCAAGCAAGTCCGAAAAAGAGCCATACGGCTTCAGATCTGAT 3780
Db 3870 CAGAAACCTTAAAGGAGTCAAGCAAGTCCGAAAAAGAGCCATACGGCTTCAGATCTGAT 3929
QY 3781 GAAACGAGTGGCTTGAAGAAAAAGGCTCAAAAGAAATATTAAGAAAAATGAAGTGA 3840
Db 3930 GAAACGAGTGGCTTGAAGAAAAAGGCTCAAAAGAAATATTAAGAAAAATGAAGTGA 3989
QY 3841 CAGAAATGTCGCGCAAAAAAGGTTAAAGAGGCGCAACCAAAACCTCTTGGTGGAGGT 3900
Db 3990 CAGAAATGTCGCGCAAAAAAGGTTAAAGAGGCGCAACCAAAACCTCTTGGTGGAGGT 4049
QY 3901 ACACCAAAAGAAAGGCAACCAATGAATACTTTTAAAGAAAGAAAGCAAAAAAATCTGCA 3960
Db 4050 ACACCAAAAGAAAGGCAACCAATGAATACTTTTAAAGAAAGAAAGCAAAAAAATCTGCA 4109
QY 3961 CCTCCAGCAGCAGAGAGAGAGAAAGAAAGAAAGCAAAAGTGAAGTACGGAACGAAG 4020
Db 4110 CCTCCAGCAGCAGAGAGAGAGAAAGAAAGAAAGCAAAAGTGAAGTACGGAACGAAG 4169
QY 4021 TCCAAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGCAGCAGAGAGCAGAAATCTCTGAA 4080
Db 4170 TCCAAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGCAGCAGAGAGCAGAAATCTCTGAA 4229

QY 4081 TCTAGTCGATTTGAATTCACACAGTCCACACAGAAAGAGGAGGAGACCATCAAAA 4140
DB 4230 TCTAGTCGATTTGAATTCACACAGTCCACACAGAAAGAGGAGGAGACCATCAAAA 4289
QY 4141 AGCCATCATCCATCAACCAACCAAAAAATGTG 4173
DB 4290 ACGCCATCATCCATCAACCAACCAAAAAATGTG 4322

RESULT 3
ABV22430
ID ABV22430 standard; cDNA; 7473 BP.
XX
XX ABV22430;
AC
XX 13-SEP-2002 (first entry)
DT
XX
XX Human prostate expression marker cDNA 22421.
DE
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX 20-FEB-2001; 2001WO-US05171.
PE
XX
XX 17-FEB-2000; 2000US-18319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-235281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI: 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PI for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX
PS Claim 1: Page 3901-3902; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
CC
XX
SQ Sequence 7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;

Query Match 99.8%; Score 4165.6; DB 23; Length 7473;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCTCATTTCAAGACTAGACCAATGATGAGAAAAATTACATATCCCGCTGGGGTCAAG 60

DB 158 ATGGCTCATTTCAAGACTAGACCAATGATGAGAAAAATTACATATCCCGCTGGGGTCAAG 217
QY 61 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 218 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 277
QY 121 ACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 278 ACTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 337
QY 181 TTACATCTTGGCTTCAGATTTTCTCAAGCATCTCTGTAAGATGTTGCTTACTGCTA 240
DB 338 TTACATCTTGGCTTCAGATTTTCTCAAGCATCTCTGTAAGATGTTGCTTACTGCTA 397
QY 241 GCGTCGCTGCTGCTGATATTTTCAAGATTTATGCTCCGGAAGCTCTTACACATCCCT 300
DB 398 GCGTCGCTGCTGCTGATATTTTCAAGATTTTATGCTCCGGAAGCTCTTACACATCCCT 457
QY 301 GATAAAGTAAAGATATATTTATGTTTATACAGACAGTTGMAAGGGCTAGAGATACA 360
DB 458 GATAAAGTAAAGATATATTTATGTTTATACAGACAGTTGMAAGGGCTAGAGATACA 517
QY 361 AAGAGCCCAATTCATAGTATTTTATTTACTTGAAGACATTGCTTGGGTCAAGTCA 420
DB 518 AAGAGCCCAATTCATAGTATTTTATTTACTTGAAGACATTGCTTGGGTCAAGTCA 577
QY 421 TATACATATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 578 TATACATATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 637
QY 481 TTATTTTCAGTTATTAACAAATGCGCCACATCAAGAAATGCAATGACATGATGATGATGAT 540
DB 638 TTATTTTCAGTTATTAACAAATGCGCCACATCAAGAAATGCAATGACATGATGATGATGAT 697
QY 541 ATGAGCTCTATTTATTTGGAAGGTGATACAGTGTCTCAGAGCTTTTGATACGTTTGA 600
DB 698 ATGAGCTCTATTTATTTGGAAGGTGATACAGTGTCTCAGAGCTTTTGATACGTTTGA 757
QY 601 GTAATCTGTTACCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
DB 758 GTAATCTGTTACCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 817
QY 661 TTACTGAAGAGACAGCTCAAGCTATTTAGGCAATTTATTTATTTATTTATTTATTTATTT 720
DB 818 TTACTGAAGAGACAGCTCAAGCTATTTAGGCAATTTATTTATTTATTTATTTATTTATTT 877
QY 721 CTGATGCTTGGGAAACATCTATCAGCGATTTGTAGAGCATGCTTTGACTTAATTTTG 780
DB 878 CTGATGCTTGGGAAACATCTATCAGCGATTTGTAGAGCATGCTTTGACTTAATTTTG 937
QY 781 GAGCTCTCAATATTTGATAGTATTTGCTGCTCTCTGTTTAAACCCAGCTTGAATTTAA 840
DB 938 GAGCTCTCAATATTTGATAGTATTTGCTGCTCTCTGTTTAAACCCAGCTTGAATTTAA 997
QY 841 TTAAGAGCAATGATATGAGAGGCGCTTACAAAGTTGTTAAACTACTGCAAAAAATGTT 900
DB 998 TTAAGAGCAATGATATGAGAGGCGCTTACAAAGTTGTTAAACTACTGCAAAAAATGTT 1057
QY 901 GGGGCAAGGATTCGAATTTGGCTTCTCAAAACCAAGCATTGGCACTGCTACTTGGGC 960
DB 1058 GGGGCAAGGATTCGAATTTGGCTTCTCAAAACCAAGCATTGGCACTGCTACTTGGGC 1117
QY 961 AGGTTATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 1118 AGGTTATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1177
QY 1021 CTCATGAACCATCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGTGAAGTCAAT 1080
DB 1178 CTCATGAACCATCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGTGAAGTCAAT 1237
QY 1081 GACCTGAGGAAGCTATTTAGACATGATTTTATTTGCTCAATATGTTACAGCTGCTTAAAG 1140

Db 1238 GACCCGTGAGGAAGCTATTAGACATGATGTATTGTGCAATAGTTACAGCTGCTAAAAAG 1297
Qy 1141 GATATCTTCGTGCAATGATCACTTACTTAATTTGTGAGAGAGAAACATTAACAAA 1200
Db 1298 GATATCTTCGTGCTAAATGATCACTTACTTAATTTGTGAGAGAGAAACATTAACAAA 1357
Qy 1201 CGATGAGAGATACGCAAGAACCCATGTATGGACTTGGCCAAATTTATPAGAAATGCT 1260
Db 1358 CGATGAGAGATACGCAAGAACCCATGTATGGACTTGGCCAAATTTATPAGAAATGCT 1417
Qy 1261 TTACAGTCAGACGTGGAAGAAAGATGCTGCAAAACAGATAGCATGATCAGAACAAATG 1320
Db 1418 TTACAGTCAGACGTGGAAGAAAGATGCTGCAAAACAGATAGCATGATCAGAACAAATG 1477
Qy 1321 CTACATATATATTAATCAAAATAGTATGTATGATGACTACTGTATGAACGGATCTTGCT 1380
Db 1478 CTACATATATATTAATCAAAATAGTATGTATGATGACTACTGTATGAACGGATCTTGCT 1537
Qy 1381 CAATPACATGTTCTCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATPACTTG 1440
Db 1538 CAATPACATGTTCTCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATPACTTG 1597
Qy 1441 TATGCGACACTGAGATTTAAATGCTGTGAAGCATTTGAATGAATGCGAAATGCTCAAAAT 1500
Db 1598 TATGCGACACTGAGATTTAAATGCTGTGAAGCATTTGAATGAATGCGAAATGCTCAAAAT 1657
Qy 1501 CTGCTCCGACATCAAGTAAGAGATTGCTGACTGTGATTAGCAACCCAAACACATGCTC 1560
Db 1658 CTGCTCCGACATCAAGTAAGAGATTGCTGACTGTGATTAGCAACCCAAACACATGCTC 1717
Qy 1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTACAGAAATTTACCTGATCCTGCT 1620
Db 1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTACAGAAATTTACCTGATCCTGCT 1777
Qy 1621 AAGGCTCAGGATTTATAGAAATTCACACAGGTGTGAAGATGATGAGAAATTAAGA 1680
Db 1778 AAGGCTCAGGATTTATAGAAATTCACACAGGTGTGAAGATGATGAGAAATTAAGA 1837
Qy 1681 AAGCAGTAGAAGTACTGTGTGATGTCACACATGCTCCTGCAACGAGCTGAAGTTGTGTG 1740
Db 1838 AAGCAGTAGAAGTACTGTGTGATGTCACACATGCTCCTGCAACGAGCTGAAGTTGTGTG 1897
Qy 1741 CGTGAATPACTAAGAGTTGGGCAACCCCAAGCCCTTACAATCTCTTCTGGAATG 1800
Db 1898 CGTGAATPACTAAGAGTTGGGCAACCCCAAGCCCTTACAATCTCTTCTGGAATG 1957
Qy 1801 ATCAAGTTCTCTTGTGAGAGATAGCACCCTGTGCACATAGATACCGAATCTATCAGTGT 1860
Db 1958 ATCAAGTTCTCTTGTGAGAGATAGCACCCTGTGCACATAGATACCGAATCTATCAGTGT 2017
Qy 1861 CTATTTAACAAGTAGAACAAATCANTAGATGAGACAGAGATGAAAGATGAGGTTGT 1920
Db 2018 CTATTTAACAAGTAGAACAAATCANTAGATGAGACAGAGATGAAAGATGAGGTTGT 2077
Qy 1921 CCAACTGTATCAAGCCATCAGAGCAGTCTTGAATGCTTAAAGTACTCTCATTTACACAT 1980
Db 2078 CCAACTGTATCAAGCCATCAGAGCAGTCTTGAATGCTTAAAGTACTCTCATTTACACAT 2137
Qy 1981 CCCATCTCATTTCTGCTGAGAAATTTGAATCATTAATGCTGCTGTGAAAAATGAT 2040
Db 2138 CCCATCTCATTTCTGCTGAGAAATTTGAATCATTAATGCTGCTGTGAAAAATGAT 2197
Qy 2041 GATGAAAAAGTAGCAAGAGCTGCACTACAATTTCAAAAAACAGGAAAGCAAAATTTGA 2100
Db 2198 GATGAAAAAGTAGCAAGAGCTGCACTACAATTTCAAAAAACAGGAAAGCAAAATTTGA 2257
Qy 2101 GAGGATTTTCCACACATCAGATCAGCTTGTCTGCTTTTACATCACAATCTAAAAA 2160
Db 2258 GAGGATTTTCCACACATCAGATCAGCTTGTCTGCTTTTACATCACAATCTAAAAA 2317
Qy 2161 GGACCCCCCGGTCAAGCAAAATATGCAATTCATTTATCCATGAGATATTTTCTAGTAA 2220
Db 2318 GGACCCCCCGGTCAAGCAAAATATGCAATTCATTTATCCATGAGATATTTTCTAGTAA 2377

Qy 2221 GAGACCCAGTTTGCACAGATATTGAGCCTGTGATTAAGAGCCTGATCAAGCAACCTG 2280
Db 2378 GAGACCCAGTTTGCACAGATATTGAGCCTGTGATTAAGAGCCTGATCAAGCAACCTG 2437
Qy 2281 GAACATCTCATTAACACATTTGTTACTATTTGGTCAATTTGCTCTCTGACCTGATCA 2340
Db 2438 GAACATCTCATTAACACATTTGTTACTATTTGGTCAATTTGCTCTCTGACCTGATCA 2497
Qy 2341 TTTGCTGCTCTTGGAAATCTTGGGTAGCTACTTTTCATTTGTAAAGATCTTCTCATGAT 2400
Db 2498 TTTGCTGCTCTTGGAAATCTTGGGTAGCTACTTTTCATTTGTAAAGATCTTCTCATGAT 2557
Qy 2401 GATCGGCTTCCAGGAAAAAGCACTAAACCTTGGGTTCCAGATGAAGATATCTCTCT 2460
Db 2558 GATCGGCTTCCAGGAAAAAGCACTAAACCTTGGGTTCCAGATGAAGATATCTCTCT 2617
Qy 2461 GAGACAAATGCTCAAAATTCAGGCTATTAAATGATGTTGATGCTCTTGGAAATGA 2520
Db 2618 GAGACAAATGCTCAAAATTCAGGCTATTAAATGATGTTGATGCTCTTGGAAATGA 2677
Qy 2521 AATATPACACATTAATCAGGAATCTTACTTAAGATTTGCTTAAACAATATTTGATAGT 2580
Db 2678 AATATPACACATTAATCAGGAATCTTACTTAAGATTTGCTTAAACAATATTTGATAGT 2737
Qy 2581 GATGAGACTTGACAGAACAGGGAAGAAATTAAGAACAGATATGTCAGCTGAGACTT 2640
Db 2738 GATGAGACTTGACAGAACAGGGAAGAAATTAAGAACAGATATGTCAGCTGAGACTT 2797
Qy 2641 GCTGCTGGAGTGTATTTGTGAAGCTGTGCACAAGAACCTGTTTACCATGAATATCACA 2700
Db 2798 GCTGCTGGAGTGTATTTGTGAAGCTGTGCACAAGAACCTGTTTACCATGAATATCACA 2857
Qy 2701 TTAGAACAATACACTATGTCATTAAGTATCAACAGATGATGCTATCAAGTAAAGACA 2760
Db 2858 TTAGAACAATACACTATGTCATTAAGTATCAACAGATGATGCTATCAAGTAAAGACA 2917
Qy 2761 GTGTTGGCCAGAAATCTTCAAAAGCCCTTCCCTTACGCTTCCACTTGAATATG 2820
Db 2918 GTGTTGGCCAGAAATCTTCAAAAGCCCTTCCCTTACGCTTCCACTTGAATATG 2977
Qy 2821 GCAATCTGTGCCCTTGTGCAAAAGATCTGTAAGAGAGAGAGCTCATGAGCA 2880
Db 2978 GCAATCTGTGCCCTTGTGCAAAAGATCTGTAAGAGAGAGAGCTCATGAGCA 3037
Qy 2881 TGTTTGGTGAATAATTAATTAAGCGGGAGTATCTGAACACAGATCAGCTGTAT 2940
Db 3038 TGTTTGGTGAATAATTAATTAAGCGGGAGTATCTGAACACAGATCAGCTGTAT 3097
Qy 2941 GAAAAATTAATGCTCTCTTCTACAGATGTTGTTGCTCATTAACAATCACCCTTTGGCA 3000
Db 3098 GAAAAATTAATGCTCTCTTCTACAGATGTTGTTGCTCATTAACAATCACCCTTTGGCA 3157
Qy 3001 CATGACCCAGATTAATGCTCAAGATGATTAAGCAACTTAAGATGTTAAAGATG 3060
Db 3158 CATGACCCAGATTAATGCTCAAGATGATTAAGCAACTTAAGATGTTAAAGATG 3217
Qy 3061 CTTTGGTTGTCTGGAATATTATGCTTAATAATGAATAAAGTACAGCTTTTATC 3120
Db 3218 CTTTGGTTGTCTGGAATATTATGCTTAATAATGAATAAAGTACAGCTTTTATC 3277
Qy 3121 AGAAAGATGCTGAAGAAATTTTAACAACAAAGATGCCCAGAGACAGATATGCAAA 3180
Db 3278 AGAAAGATGCTGAAGAAATTTTAACAACAAAGATGCCCAGAGACAGATATGCAAA 3337
Qy 3181 ATGAATGAAGAAATCTGTACAGTGTGTGATGTTGCCATGAATATCATCATGCAAGAT 3240
Db 3338 ATGAATGAAGAAATCTGTACAGTGTGTGATGTTGCCATGAATATCATCATGCAAGAT 3397
Qy 3241 ACTACATACAGTTTGAATCTCTTAAGAACCCGCTGACTACAGCTGCTTCTTCACTCA 3300
Db 3398 ACTACATACAGTTTGAATCTCTTAAGAACCCGCTGACTACAGCTGCTTCTTCACTCA 3457

QY 3301 CCTGACAGAAATTTCACTAACACCCAAAATATCTGCTCTCTGAAATGAATCATTTTTC 3360
|||
DB 3458 CCTGACAGAAATTTCACTAACACCCAAAATATCTGCTCTCTGAAATGAATCATTTTTC 3517
QY 3361 ACTCTGGAAAACCTTAAACACCAATGTTCTAGAGCTGTTAAACAAGCCATTTCATCA 3420
|||
DB 3518 ACTCTGGAAAACCTTAAACACCAATGTTCTAGAGCTGTTAAACAAGCCATTTCATCA 3577
QY 3421 GCAGCAGCAATCTCAGACCAATCATCAGCAATGCAATGCAATGCAATGCAATGCAAGC 3480
|||
DB 3578 GCAGCAGCAATCTCAGACCAATCATCAGCAATGCAATGCAATGCAATGCAATGCAAGC 3637
QY 3481 AGCTCAATCCAACTCTCTCTGGAAGATTAAGGGGAGGCTTGAAGTTCTGAAATGAT 3540
|||
DB 3638 AGCTCAATCCAACTCTCTCTGGAAGATTAAGGGGAGGCTTGAAGTTCTGAAATGAT 3697
QY 3541 CACGTGAAATGAAGTTTACCAATGTTCTCACCCTTTCGCGGGGAAAAAGACACAG 3600
|||
DB 3698 CACGTGAAATGAAGTTTACCAATGTTCTCACCCTTTCGCGGGGAAAAAGACACAG 3757
QY 3601 AGACAGCACTCTGATCTTGTAAAGTCTGAATTTGAGAAGCCTAGAGGAGGAAAAAAG 3660
|||
DB 3758 AGACAGCACTCTGATCTTGTAAAGTCTGAATTTGAGAAGCCTAGAGGAGGAAAAAAGC 3817
QY 3661 CCCGTCAAGAAAGAGGAGGAAATTAAGTATGATGACTTGAATAGTTGTTGACAGAA 3720
|||
DB 3818 CCCGTCAAGAAAGAGGAGGAAATTAAGTATGATGACTTGAATAGTTGTTGACAGAA 3877
QY 3721 CAGAAACCTTAAAGGCACTCAGCCGAAAGTCGAAAGAGGCTTACGGCTTCGAATCTGAT 3780
|||
DB 3878 CAGAAACCTTAAAGGCACTCAGCCGAAAGTCGAAAGAGGCTTACGGCTTCGAATCTGAT 3937
QY 3781 GAACAGAGTGGCTGAGGAAAAAGAGGCTTCAAGAATATATTAGAAATGAAGATGAA 3840
|||
DB 3938 GAACAGAGTGGCTGAGGAAAAAGAGGCTTCAAGAATATATTAGAAATGAAGATGAA 3997
QY 3841 CAGAAATGCTCCGCAAAAAAGGTTAAAGAGGCGGACCAACCACTCTTGTGTGAGGT 3900
|||
DB 3998 CAGAAATGCTCCGCAAAAAAGGTTAAAGAGGCGGACCAACCACTCTTGTGTGAGGT 4057
QY 3901 ACACCAAAAGAAAGAGCCCAACATGAATCTTAAAAAGAGACCAAAAAAATCTGGA 3960
|||
DB 4058 ACACCAAAAGAAAGAGCCCAACATGAATCTTAAAAAGAGACCAAAAAAATCTGGA 4117
QY 3961 CCTCCAGCACCAGAGGAGGAGGAAAGAAAGAAAGAAAGTGAATATCGGAAACAGAG 4020
|||
DB 4118 CCTCCAGCACCAGAGGAGGAGGAAAGAAAGAAAGAAAGTGAATATCGGAAACAGAG 4177
QY 4021 TCCAAAAGCAACAGCAGCGAGTGTCAAGGAGACACAGCAGAGACGAATCTCTGAA 4080
|||
DB 4178 TCCAAAAGCAACAGCAGCGAGTGTCAAGGAGACACAGCAGAGACGAATCTCTGAA 4237
QY 4081 TCTAGTCAATTTGAATCCACAGAGTCCACACCAAGAAAGAGAGCAATCTCAAAA 4140
|||
DB 4238 TCTAGTCAATTTGAATCCACAGAGTCCACACCAAGAAAGAGAGCAATCTCAAAA 4297
QY 4141 ACGCCATCACCATCAACAACCAAAAAAATGCT 4172
|||
DB 4298 ACGCCATCACCATCAACAACCAAAAAAATGCT 4329
|||
RESULT 4
ABV25469
ID ABV25469 standard; cdna: 7473 BP.
XX
AC ABV25469;
XX
DF 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cdna 25460.
XX
KW Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene: ss.

XX XX Homo sapiens.
OS |||
XX |||
PN WO200160860-A2.
XX |||
XX 23-AUG-2001.
XX |||
PF 20-FEB-2001; 2001WO-US05171.
XX |||
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX |||
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX |||
XX Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
XX |||
DR Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
PT Claim 1; Page 5055-5057; 11750pp; English.
PS |||
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;
XX
Query Match 99.8%; Score 4164; DB 23; Length 7473;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGCTCATTCMAAGACTAGACCAATGATGAAAAATTAATATTCGCGCTGGGGTCAAG 60
|||
DB 158 ATGGCTCATTCMAAGACTAGACCAATGATGAAAAATTAATATTCGCGCTGGGGTCAAG 217
QY 61 GAATATTCACATTAATAATATCTAAAGAGACAGATGTTGAGACGATTAAAGATGTTGTAAA 120
|||
DB 218 GAATATTCACATTAATAATATCTAAAGAGAGATGTTGAGACGATTAAAGATGTTGTAAA 277
|||
QY 121 ACTTTATGATGATGAGACAGAGCTGGAAGAAAGAAAGAGCTTTATTTAAACCTAGCT 180
|||
DB 278 ACTTTATGATGATGAGACAGAGCTGGAAGAAAGAAAGAGCTTTATTTAAACCTAGCT 337
|||
QY 181 TTACATCTTGCTTCAGATTTTCTCTCAAGCATCTGGTAAAGATGTTGCTTACTGATA 240
|||
DB 338 TTACATCTTGCTTCAGATTTTCTCTCAAGCATCTGGTAAAGATGTTGCTTACTGATA 397
|||
QY 241 GCTTGCTGCTTGGTGAATTTTCAAGATTTATGCTCTGAAAGCTCCTTACACATCCCT 300
|||
DB 398 GCTTGCTGCTTGGTGAATTTTCAAGATTTATGCTCTGAAAGCTCCTTACACATCCCT 457
|||
QY 301 GATAAAGTAAAGATATATTTATTTATTAACAAGACAGTTGAAGGGGCTAGAGATACA 360
|||
DB 458 GATAAAGTAAAGATATATTTATTTATTAACAAGACAGTTGAAGGGGCTAGAGATACA 517
|||

QY	361	AAGAGCCACAAATTCATAGGTATTTTATTACTTGAGAACATTGCTTGCGTCAATGCA	420
Db	518	AAGAGCCACCAATTCATAGGTATTTTATTACTTGAGAACATTGCTTGCGTCAATGCA	577
QY	421	TATTAACATATGCTTTGAGTGTAGAGATGTGCAATGAATTTTCCACCACCTATACGAAC	480
Db	578	TATTAACATATGCTTTGAGTGTAGAGATGTGCAATGAATTTTCCACCACCTATACGAAC	637
QY	481	TTAATTTTCAGTTATTAACCAATGGCCACAAATCAGAAAGTCCATATGACATGGTAGACCTT	540
Db	638	TTAATTTTCAGTTATTAACCAATGGCCACAAATCAGAAAGTCCATATGACATGGTAGACCTT	697
Db	541	ATGAGCCTCATATTTTGTGAAGGTGATACAGTGTCTAGAGACCTTTTGGATACGGTTTAA	600
QY	698	ATGAGCCTCATATTTTGTGAAGGTGATACAGTGTCTAGAGACCTTTTGGATACGGTTTAA	757
Db	601	GTAATATCGTACCTGCTCATAGATTTTAAACAAGCAGATATGATTTGGCAAGAGCT	660
QY	758	GTAATATCGTACCTGCTCATAGATTTTAAACAAGCAGATATGATTTGGCAAGAGCT	817
Db	661	TTACTGAAGAGCAGCTCAAGCTATTAGCCATATATTACACCTTTTTTATACAGTT	720
QY	818	TTACTGAAGAGCAGCTCAAGCTATTAGCCATATATTACACCTTTTTTATACAGTT	877
Db	721	CTGATGCTTGGGAAAACATCATATACAGCATTTGTGAGAGCATGTCTTGACTTATTTTG	780
QY	878	CTGATGCTTGGGAAAACATCATATACAGCATTTGTGAGAGCATGTCTTGACTTATTTTG	937
Db	781	GAGCTCTACAAATATGTAGATGTCATTTGCTGCTCTCTGTTTAAACCACCTGAATTTTAA	840
QY	938	GAGCTCTACAAATATGTAGATGTCATTTGCTGCTCTCTGTTTAAACCACCTGAATTTTAA	997
Db	841	TTTAAAGCAGCATGATTAATGAGAGAGCGCTACAAAGTGTGTTAACTACTGGCAAAATGTTT	900
QY	998	TTTAAAGCAGCATGATTAATGAGAGAGCGCTACAAAGTGTGTTAACTACTGGCAAAATGTTT	1057
Db	901	GGGGCAAAAGATTTAGAAATTTGGCTTCTCAAAAACAAGCCATTGGCAGTGCATCTTGAGG	960
QY	1058	GGGGCAAAAGATTTAGAAATTTGGCTTCTCAAAAACAAGCCATTGGCAGTGCATCTTGAGG	1112
Db	961	AGGTTTAATGATATTCATGTACCAATCCGCTCGAGATGTGTGAATTTGCTAGCCATTGT	1020
QY	1118	AGGTTTAATGATATTCATGTACCAATCCGCTCGAGATGTGTGAATTTGCTAGCCATTGT	1177
Db	1021	CTCATGAAACCATCCGAGATTTAGCAAAACACTTAACAGATATCTTAAAGTAGAGTCACAT	1088
QY	1178	CTCATGAAACCATCCGAGATTTAGCAAAACACTTAACAGATATCTTAAAGTAGAGTCACAT	1237
Db	1081	GACCCCTGAGAGAGCTATTAGACATGATGTTATTTGTCTCAATAGTTACAGCTGCTAAAAG	1140
QY	1238	GACCCCTGAGAGAGCTATTAGACATGATGTTATTTGTCTCAATAGTTACAGCTGCTAAAAG	1297
Db	1141	GATATTTCTCTGCGTCAATGATCCTACTTACTTATTTTGTGAGAGAGAGAACTTTAGACAAA	1207
QY	1298	GATATTTCTCTGCGTCAATGATCCTACTTACTTATTTTGTGAGAGAGAGAACTTTAGACAAA	1357
Db	1201	CGATGAGAGATGACCAAAAGAGCCATGATGGAGCTTGCCCAAAATTTTATAGAAATATGCT	1266
QY	1358	CGATGAGAGATGACCAAAAGAGCCATGATGGAGCTTGCCCAAAATTTTATAGAAATATGCT	1411
Db	1261	TTTACAGTCAGCAGCTGCAAAAGATGCTGCAAAACAGATAGCATGATCAAAAGACAAATTG	1320
QY	1418	TTTACAGTCAGCAGCTGCAAAAGATGCTGCAAAACAGATAGCATGATCAAAAGACAAATTG	1477
Db	1321	CTACATATATTTTATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1380
QY	1478	CTACATATATTTTATCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1537
Db	1381	CAATACATGCTTCTCCACAAATTTAGAAACTACAGAAAGGATGAATGCTTATATTACTTG	1440
QY	1538	CAATACATGCTTCTCCACAAATTTAGAAACTACAGAAAGGATGAATGCTTATATTACTTG	1597

OY	1441	TATGCCACAC	TGGATTTAAATGCTGTGTAAGACATTTGAATGAAATGTGGAATGTCAAAAT	1500
Db	1598	TATCCACAC	ATCGGATTTAAATGCTGTGTAAGACATTTGAATGAAATGTGGAATGTCAAAAT	1557
OY	1501	CTGCTCCGAC	ATCAAGTAAGAGATTTGGCTTACTTGATTTAAGCAACCCAAACAGATGCC	1560
Db	1658	CTGCTCCGAC	ATCAAGTAAGAGATTTGGCTTACTTGATTTAAGCAACCCAAACAGATGCC	1717
OY	1561	AGTGTCAAG	CCCATATTTTCAAAAAGTGTGTTATTAACAAGAAATTTTACCTGATCCTGCT	1620
Db	1718	AGTGTCAAG	CCCATATTTTCAAAAAGTGTGTTATTAACAAGAAATTTTACCTGATCCTGCT	1777
OY	1621	AAGCCTACGA	CTTTTCATGAGAAATTTACACACAGCTGTTAGACAGATGATGAGAAATTAAGA	1680
Db	1778	AAGCCTACGA	CTTTTCATGAGAAATTTACACACAGCTGTTAGACAGATGATGAGAAATTAAGA	1837
OY	1681	AAGCAGTTAG	AAGTACTTGTATGATCAACAATGCTCTCGCAAGCAGCGCTGAAGGTGTGCTG	1740
Db	1838	AAGCAGTTAG	AAGTACTTGTATGATCAACAATGCTCTCGCAAGCAGCGCTGAAGGTGTGCTG	1897
OY	1741	CGTGAATTA	CTAAGAAGTTGGGCAACCCCAACAGCCTACAAATCTTTCTCGGAATG	1800
Db	1898	CGTGAATTA	CTAAGAAGTTGGGCAACCCCAACAGCCTACAAATCTTTCTCGGAATG	1957
OY	1801	ATCAAGTTT	CTCTTGGAGAGATATGACACTGTGTGCATATGATACCGAATCTATCAGTCT	1860
Db	1958	ATCAAGTTT	CTCTTGGAGAGATATGACACTGTGTGCATATGATACCGAATCTATCAGTCT	2017
OY	1861	CTTATTTAA	CAAGTGAACAAATTCAAATAGTGGAAACACAGATGATGAAAGTGAAGGTGCT	1920
Db	2018	CTTATTTAA	CAAGTGAACAAATTCAAATAGTGGAAACACAGATGATGAAAGTGAAGGTGCT	2077
OY	1921	CCAACGTAT	CAAGCCTACAGACAGAGCTTGAACCTGTTAAAGTACTCTCATTTACACAT	1980
Db	2078	CCAACGTAT	CAAGCCTACAGACAGAGCTTGAACCTGTTAAAGTACTCTCATTTACACAT	2137
OY	1981	CCCATCTCA	TTTCATTTCTGCTGTAACAACTTTGAATCATTTACTGCGCTGTCTGAAAAATGAT	2040
Db	2138	CCCATCTCA	TTTCATTTCTGCTGTAACAACTTTGAATCATTTACTGCGCTGTCTGAAAAATGAT	2197
OY	2041	GATCAAAAA	AGTAGCAGAACTGCACTACAAATTTTCAAAAACACAGAGCAAAATTTGAA	2100
Db	2198	GATCAAAAA	AGTAGCAGAACTGCACTACAAATTTTCAAAAACACAGAGCAAAATTTGAA	2257
OY	2101	GAGGATTTT	CCACACATCAGATTCAGCCTGCTTCCCTTTTACATCAACAAATCTAAAAAA	2160
Db	2258	GAGGATTTT	CCACACATCAGATTCAGCCTGCTTCCCTTTTACATCAACAAATCTAAAAAA	2317
OY	2161	GGACCCCC	CCGTCACAGCCAAATATGCCATTCATTTGATTCATGCGATGAAATTTTCTAGTAAA	2220
Db	2318	GGACCCCC	CCGTCACAGCCAAATATGCCATTCATTTGATTCATGCGATGAAATTTTCTAGTAAA	2377
OY	2221	GAGACCCA	GTTTGCACAGATATTTGAGCCTTGCATTAAGCCTAGATTCACAGACCTG	2280
Db	2378	GAGACCCA	GTTTGCACAGATATTTGAGCCTTGCATTAAGCCTAGATTCACAGACCTG	2437
OY	2281	GAAATATCT	CAATTAACCATTTGGTATATTTGCTTCCCTTGCACCCGATGACAA	2340
Db	2438	GAAATATCT	CAATTAACCATTTGGTATATTTGCTTCCCTTGCACCCGATGACAA	2497
OY	2341	TTTGTCTCT	CTCTTGAATCTTGGGTAGCTACTTTCATTTGGAAGAATCTTCTCATGAAT	2400
Db	2498	TTTGTCTCT	CTCTTGAATCTTGGGTAGCTACTTTCATTTGGAAGAATCTTCTCATGAAT	2557
OY	2401	GATCGGCTT	CCAGGAAAAAGACAACCTTAACTTTGGGTTCCAGATGAGAAGATATCTCT	2460
Db	2558	GATCGGCTT	CCAGGAAAAAGACAACCTTAACTTTGGGTTCCAGATGAGAAGATATCTCT	2617
OY	2461	GAGACATGT	GTCACAAATTTACAGCTATTTAAATGATGTTGGATGGCTACTTGGAAATGAAA	2520
Db	2618	GAGACATGT	GTCACAAATTTACAGCTATTTAAATGATGTTGGATGGCTACTTGGAAATGAAA	2677
OY	2521	AATTAATCA	AGTAATAGGAACCTTACCTTAAGATTTGCTTAACACAAATATTTGCATAGT	2580

```

Db 2678 AATATACAGTAATACAGAACTCTACCTTAAGATGTACAAACAATATGATAGT 2737
Qy 2581 GATGAGAGCTTGACAGAAAGGGGAAATTTAGTAACAGAAATATCAGCTGAGCT 2640
Db 2738 GATGAGAGCTTGACAGAAAGGGGAAATTTAGTAACAGAAATATCAGCTGAGCT 2797
Qy 2641 GCTGCTGGAGTGTATTTGTAAGCTGACACAGAACCCCTTTACATGAATATCATCA 2700
Db 2798 GCTGCTGGAGTGTATTTGTAAGCTGACACAGAACCCCTTTACATGAATATCATCA 2857
Qy 2701 TTAAGACAAATATCAGCTATGTGCAATACATCAATGTAATGCTATCAAGTAA 2760
Db 2858 TTAAGACAAATATCAGCTATGTGCAATACATCAATGTAATGCTATCAAGTAA 2917
Qy 2761 GTGTTTGGCCAGAACTTACAAAGGCTTCCGCTTACGGCTTCCACTGAGATATG 2820
Db 2918 GTGTTTGGCCAGAACTTACAAAGGCTTCCGCTTACGGCTTCCACTGAGATATG 2977
Qy 2821 GCAATCTGTGCTTGTGCAAAAGATCCTGTAAAGAGAGAGAGAGATGCTATGCGCA 2880
Db 2978 GCAATCTGTGCTTGTGCAAAAGATCCTGTAAAGAGAGAGAGATGCTATGCGCA 3037
Qy 2881 TGTTTGGTGAATAATATTAATGTAAAGCGGGAGATATCTGAAGCAGCATGCACTGTATG 2940
Db 3038 TGTTTGGTGAATAATATTAATGTAAAGCGGGAGATATCTGAAGCAGCATGCACTGTATG 3097
Qy 2941 GAAATAATATGTCTCTACAGAGATATGTGTCATATACAAATCACCTTTGGCA 3000
Db 3098 GAAATAATATGTCTCTCTACAGAGATATGTGTCATATACAAATCACCTTTGGCA 3157
Qy 3001 CATACCCAGATATATGTCAAAGTACAGGATATGGAACAATTAAAGATTTAAAGATGT 3060
Db 3158 CATACCCAGATATATGTCAAAGTACAGGATATGGAACAATTAAAGATTTAAAGATGT 3217
Qy 3061 CTTTGGTTTGTTCGAAATATTAATGCTTAAATGAAATPAACAGTCAGCTTTATC 3120
Db 3218 CTTTGGTTTGTTCGAAATATTAATGCTTAAATGAAATPAACAGTCAGCTTTATC 3277
Qy 3121 AGAAGATGTAGAAATATTAACAAACAAAGATGCCCCAAGACAGATGATGCAAAA 3180
Db 3278 AGAAGATGTAGAAATATTAACAAACAAAGATGCCCCAAGACAGATGATGCAAAA 3337
Qy 3181 ATGATGAAAACTGTACACTGTGTGATGTGTCATGAATATCATATGTCAAAGAGT 3240
Db 3338 ATGATGAAAACTGTACACTGTGTGATGTGTCATGAATATCATATGTCAAAGAGT 3397
Qy 3241 ACTACATPACGTTTGGAAATCCTTAAGACCGGACTACAGCTCGTTCTCTCA 3300
Db 3398 ACTACATPACGTTTGGAAATCCTTAAGACCGGACTACAGCTCGTTCTCTCA 3457
Qy 3301 CCTGACAGAAATTTAGTAACCAAAATTTATGCTCCTCGAATGAATCATTTTTC 3360
Db 3458 CCTGACAGAAATTTAGTAACCAAAATTTATGCTCCTCGAATGAATCATTTTTC 3517
Qy 3361 ACTCCTGGAAAACTAAAAACAACAAATGTTAGAGCTGTTAACAGCCACTTCATCA 3420
Db 3518 ACTCCTGGAAAACTAAAAACAACAAATGTTAGAGCTGTTAACAGCCACTTCATCA 3577
Qy 3421 GCAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAATGTAAGAAATGCAAGCAGC 3480
Db 3578 GCAGGCAAGCAATCTCAGACCAATCATCAGAAATGGAATGTAAGAAATGCAAGCAGC 3637
Qy 3481 AGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGGGAGGCTTGATATGTTGAAATGAT 3540
Db 3638 AGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGGGAGGCTTGATATGTTGAAATGAT 3697
Qy 3541 CACAGTGAATAATGAAGATTACACAATGTCTTACCTTCCGCGGGAAAAAGTGACAG 3600
Db 3698 CACAGTGAATAATGAAGATTACACAATGTCTTACCTTCCGCGGGAAAAAGTGACAG 3757
Qy 3601 AGAGACGACTGTATCTTGAAGGTCTGAATTTGGAAGCCTAGAGGAGGAAAAAACG 3660

```

```

Db 3758 AGAGACGACTGTATCTTGAAGGTCTGAATTTGGAAGACCTAGAGCAGGAAAAAACG 3817
Qy 3661 CCCGTCACAGAAACAGAGAGAAATTTAGTATGATGATGACTGACTACATAGTGGTACAGAA 3720
Db 3818 CCCGTCACAGAAACAGAGAGAAATTTAGTATGATGATGACTGACTACATAGTGGTACAGAA 3877
Qy 3721 CAGAAACCTTAAGGAGTACAGAGAAATGCGAAAGAGGCGCTTACGCTTCAAGATCTGAT 3780
Db 3878 CAGAAACCTTAAGGAGTACAGAGAAATGCGAAAGAGGCGCTTACGCTTCAAGATCTGAT 3937
Qy 3781 GAACAGCAGTGGCTTGAGGAAAAAGGCTCAAGAGATATTTTGAATAATGAGATGAA 3840
Db 3938 GAACAGCAGTGGCTTGAGGAAAAAGGCTCAAGAGATATTTTGAATAATGAGATGAA 3997
Qy 3841 CAGAAATGCTCCGCCCCAAAAAGGTTAAAGAGGCGCAGCACCAAAACCTTGTGTGAGGT 3900
Db 3998 CAGAAATGCTCCGCCCCAAAAAGGTTAAAGAGGCGCAGCACCAAAACCTTGTGTGAGGT 4057
Qy 3901 ACACCAAAAGAAAGACCAACATGAAACTTCTAATAAAGAGAGCAAAAAAATCTGGA 3960
Db 4058 ACACCAAAAGAAAGACCAACATGAAACTTCTAATAAAGAGAGCAAAAAAATCTGGA 4117
Qy 3961 CCTCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 4118 CCTCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4177
Qy 4021 TCCAAAAGCAAAACAGCAGCAGTGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4178 TCCAAAAGCAAAACAGCAGCAGTGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4237
Qy 4081 TCTAGTGAATTTGAATTCACACAGTTCACACCACAGAAAGAGAGAGAGAGAGAGAGAG 4140
Db 4238 TCTAGTGAATTTGAATTCACACAGTTCACACCACAGAAAGAGAGAGAGAGAGAGAGAG 4297
Qy 4141 AGCCATCACCATCCACACCAACCAAAAAAATATG 4172
Db 4298 AGCCATCACCATCCACACCAACCAAAAAAATATG 4329

RESULT 5
ABV28244
ID ABV28244 standard; cDNA; 7473 BP.
XX ABV28244;
AC 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 28235.
XX
KW Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
WO WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-18319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-235281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX

```

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1, Page 5869-5870; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 7473 BP; 2458 A; 1244 C; 1479 G; 2285 T; 7 other:

Query Match 99.8%; Score 4164; DB 23; Length 7473;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4167; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAGGCTCATTCAAAGCTAGGACCAATGATGAGAAAATTCATATCCGCTGGGGTCAAG 60
DB 158 AAGGCTCATTCAAAGCTAGGACCAATGATGAGAAAATTCATATCCGCTGGGGTCAAG 217

QY 61 GAAATATCGATTAATATCTAAGAGAGATGCTAGACGATTAAGATGCTTGTGAAA 120
DB 218 GAAATATCGATTAATATCTAAGAGAGATGCTAGACGATTAAGATGCTTGTGAAA 277

QY 121 ACTTTATGATATGAGACGACGACTGTGAAGAGAAAGAGCTTATTTAACTAGCT 180
DB 278 ACTTTATGATATGAGACGACGACTGTGAAGAGAAAGAGCTTATTTAACTAGCT 337

QY 181 TTACATCTTGCCTTCAGATTTTCTTCAAGCATCTGCTGTAAGATGCTTGCCTTACGTGA 240
DB 338 TTACATCTTGCCTTCAGATTTTCTTCAAGCATCTGCTGTAAGATGCTTGCCTTACGTGA 397

QY 241 GCCTGCTGCTGCTGCTATATTTTTCAGAGTTTATGCTGCTGAAGCTGCTTACACATCCCT 300
DB 398 GCCTGCTGCTGCTGCTATATTTTTCAGAGTTTATGCTGCTGAAGCTGCTTACACATCCCT 457

QY 301 GATTAACCTAAGAGATATATTTATTTATTAACAAGACAGTTGAAGGGCTAGAGATACA 360
DB 458 GATTAACCTAAGAGATATATTTATTTATTAACAAGACAGTTGAAGGGCTAGAGATACA 517

QY 361 AAGAGCCCAATTCATAGATTTTATTTACTTGAGAACCTTCTGCTGGCTCAAGTCA 420
DB 518 AAGAGCCCAATTCATAGATTTTATTTTACTTGAGAACCTTCTGCTGGCTCAAGTCA 577

QY 421 TATTAACATATGCTTGAAGTATAGATACCAATGAATTTTCCAGAGCTATACAGAAC 480
DB 578 TATTAACATATGCTTGAAGTATAGATACCAATGAATTTTCCAGAGCTATACAGAAC 637

QY 481 TATATTTTCAGTTATTAACAAGTGGCCCAATCAGAAAGTCCATATGACATGTGTAGACTT 540
DB 638 TATATTTTCAGTTATTAACAAGTGGCCCAATCAGAAAGTCCATATGACATGTGTAGACTT 597

QY 541 AAGACCTCTATATTTTGAAGGTGATATAGTGTCTCAGAGCTTTTGATACGTTTAA 600
DB 698 AAGACCTCTATATTTTGAAGGTGATATAGTGTCTCAGAGCTTTTGATACGTTTAA 757

QY 601 GTAATATCTGTAAGTCTCATTAAGATTTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
DB 758 GTAATATCTGTAAGTCTCATTAAGATTTTAAACAAGCAAGCATATGATTTGGCAAGGCT 817

QY 661 TTACTGAAGAGACGCTCAAGCTATTTGAGCATATATTTACCACTTTTATATACAGTT 720

DB 818 TTACTGAAGAGACGCTCAAGCTATTTGAGCATATATTTACCACTTTTATATACAGTT 877

QY 721 CTGATCTGTGGGAAACATCTATACAGCATTTGTGCAGAGATGCTTTGACTTAATTTTG 780

DB 878 CTGATCTGTGGGAAACATCTATACAGCATTTGTGCAGAGATGCTTTGACTTAATTTTG 937

QY 781 GAGCTCTACAAATTTATAGATATGTCATTTGCTGCTGCTGTGTATACCAGCTTGAATTTAA 840

DB 938 GAGCTCTACAAATTTATAGATATGTCATTTGCTGCTGCTGTGTATACCAGCTTGAATTTAA 997

QY 841 TTAAGAGCAATGATATAGAGAGCGCTTACAAAGTTTAACTACTGGCAAAATGTTT 900

DB 998 TTAAGAGCAATGATATAGAGAGCGCTTACAAAGTTTAACTACTGGCAAAATGTTT 1057

QY 901 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAACAGCCACTTTGGCAGTCTCTTGGG 960

DB 1058 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAACAGCCACTTTGGCAGTCTCTTGGG 1117

QY 961 AGGTTTAATGATATCCATGATACCAATCCGCTGGAATGTGTAAATTTGCTAGCCATTTG 1020

DB 1118 AGGTTTAATGATATCCATGATACCAATCCGCTGGAATGTGTAAATTTGCTAGCCATTTG 1177

QY 1021 CTCATGAACCATCTGATTTTACGAAAGACTTAAAGAGATCTTAAAGTGAAGTCAAT 1080

DB 1178 CTCATGAACCATCTGATTTTACGAAAGACTTAAAGAGATCTTAAAGTGAAGTCAAT 1237

QY 1081 GACCTTGAGAGACCTTTTACATGATTTATTTGTCTCAATAGTTACAGCTCTTAAAG 1140

DB 1238 GACCTTGAGAGACCTTTTACATGATTTATTTGTCTCAATAGTTACAGCTCTTAAAG 1297

QY 1141 GATATCTTCTGTCTCAATGATCTACTTAATTTTGTGAGAGAGAAACATATGACAA 1200

DB 1298 GATATCTTCTGTCTCAATGATCTACTTAATTTTGTGAGAGAGAAACATATGACAA 1357

QY 1201 CGATGAGAGATACGAAAGAGCCATGATGGAATTTGCCAAATTTTAAAGAAATATGCT 1260

DB 1358 CGATGAGAGATACGAAAGAGCCATGATGGAATTTGCCAAATTTTAAAGAAATATGCT 1417

QY 1261 TTACATCTGAGACCTGGAAGAGTGTGCAAAACAGATAGATGATCAAGCAAAATTTG 1320

DB 1418 TTACATCTGAGACCTGGAAGAGTGTGCAAAACAGATAGATGATCAAGCAAAATTTG 1477

QY 1321 CTACATATATATATCAAAATATATGATGATGATGATGATGATGATGATGATGATGAT 1380

DB 1478 CTACATATATATATCAAAATATATGATGATGATGATGATGATGATGATGATGATGAT 1537

QY 1381 CAATACATGCTTCTCAATATTTAGAACTACAGAAACGATGAATGCTTATATTTACTTG 1440

DB 1538 CAATACATGCTTCTCAATATTTAGAACTACAGAAACGATGAATGCTTATATTTACTTG 1597

QY 1441 TATGCCACACTGATTTTAAATGCTGTGAAGACCTTGAATGAATGTGGAATGTCAAAAT 1500

DB 1598 TATGCCACACTGATTTTAAATGCTGTGAAGACCTTGAATGAATGTGGAATGTCAAAAT 1657

QY 1501 CTGCTCCGACATCAAGTAAAGATTTGCTTGAATTTTAAAGCAACCCAAACAGATGCT 1560

DB 1658 CTGCTCCGACATCAAGTAAAGATTTGCTTGAATTTTAAAGCAACCCAAACAGATGCT 1717

QY 1561 AGTGTCAAGCCATATTTTCAAAAGTGTATTTATTAAGAAATTTTACCTGATCTGCT 1620

DB 1718 AGTGTCAAGCCATATTTTCAAAAGTGTATTTATTAAGAAATTTTACCTGATCTGCT 1777

QY 1621 AAGGCTCAGATTTTCAATGAAGAAATTCACACAGGTTTGAAGATATATGAAGAAATTA 1680

DB 1778 AAGGCTCAGATTTTCAATGAAGAAATTCACACAGGTTTGAAGATATATGAAGAAATTA 1837

QY 1681 AAGCATTTGAAGTACTTGTATGTCACAAATGCTCTGCAAGAGGCTAAAGTTGTGTG 1740

DB 1838 AAGCATTTGAAGTACTTGTATGTCACAAATGCTCTGCAAGAGGCTAAAGTTGTGTG 1897

QY 1741 CGTGAATTAAGTAAAGATTTGGCAACCCCAACAGCTTACAAATCTTTTCTGGAATG 1800

Db	1898	CGTGAATAACTAGAACTTGGGCAACCCCAACAGCCTTACAAATCCTTCTCGGAATG	1957
Qy	1801	ATCAGAGTTTCCTTGGAGAGATAGACACCTGTGCACATAGATACGAATCTATCAGTCT	1866
Db	1958	ATCAAGTTCTCTTGGAGAGATAGCACCTGTGCACATAGATACGAATCTATCAGTCT	2017
Qy	1861	CTTATTTAAACAGTGAACAAATCAATAGATGGAAACAGCATGATGAAGATGAGGTGT	1920
Db	2018	CTTATTTAAACAGTGAACAAATCAATAGATGGAAACAGCATGATGAAGATGAGGTGT	2077
Qy	1921	CCAATGTATCAAGCCATAGAGACAGGTCTTAACGCTTAAGATACCTCATTTACACAT	1980
Db	2078	CCACTGTATCAAGCCATAGAGACAGGTCTTAACGCTTAAGATACCTCATTTACACAT	2137
Qy	1981	CCCATCTCATTTCAATCTGCTGAACACATTTGAATCATCTACGTGGTGTCTCAAAATGGAT	2040
Db	2138	CCCATCTCATTTCAATCTGCTGAACACATTTGAATCATCTACGTGGTGTCTCAAAATGGAT	2197
Qy	2041	GATGAAAAAGTGAAGAACTGCGACTACAAATTTTCAAAAACACAGAAACAAATTGAA	2100
Db	2198	GATGAAAAAGTGAAGAACTGCGACTACAAATTTTCAAAAACACAGAAACAAATTGAA	2257
Qy	2101	GAGATTTTCCACACATAGATACAGATCGAGCCTGCTCTCTGTTTACATCAACAAATCTAAAAA	2166
Db	2258	GAGATTTTCCACACATAGATACAGATCGAGCCTGCTCTCTGTTTACATCAACAAATCTAAAAA	2317
Qy	2161	GGACCCCCCGCTCAAGCCCAAAATATGCCATTCATTTGATATCCATGGCATATTTTCTAGTAAA	2220
Db	2318	GGACCCCCCGCTCAAGCCCAAAATATGCCATTCATTTGATATCCATGGCATATTTTCTAGTAAA	2377
Qy	2221	GAGACCCAGTTTGCACAGATATTTGAGCCTGTGCATTAAGACCTTAGATCCAAAGAACCTG	2280
Db	2378	GAGACCCAGTTTGCACAGATATTTGAGCCTGTGCATTAAGACCTTAGATCCAAAGAACCTG	2437
Qy	2281	GAACTATTCATTAACACCATTTGGTTACTTTGGTATATTTGGTCTTCCTTGCAACCTGATCAA	2340
Db	2438	GAACTATTCATTAACACCATTTGGTTACTTTGGTATATTTGGTCTTCCTTGCAACCTGATCAA	2497
Qy	2341	TTTGTGCTCCTTGGAAATCTTGGGTACTTCTTTCATTTGAAAGATCTTCTCATGAT	2400
Db	2498	TTTGTGCTCCTTGGAAATCTTGGGTACTTCTTTCATTTGAAAGATCTTCTCATGAT	2557
Qy	2401	GATGCGCTTCCAGGGAAAAAGACACTAACTTTGGGTTCCAGATGAAGAATATCTCT	2460
Db	2558	GATGCGCTTCCAGGGAAAAAGACACTAACTTTGGGTTCCAGATGAAGAATATCTCT	2617
Qy	2461	GAGACAATGGTCAAAATTCAGGCTATTTAAATGATAGTTCGATGGCTACTTGGAAATGAAA	2520
Db	2618	GAGACAATGGTCAAAATTCAGGCTATTTAAATGATAGTTCGATGGCTACTTGGAAATGAAA	2677
Qy	2521	AATAATTCACAGTAAATTCAGGAACCTCTACCTTTAAGATTGCTATACACAACAAATTTGCATAGT	2580
Db	2678	AATAATTCACAGTAAATTCAGGAACCTCTACCTTTAAGATTGCTATACACAACAAATTTGCATAGT	2737
Qy	2581	GATGAGACTTGCAGAAACAGGGGAAAAATTAGTAAACAGATATGTCACTGTAGACTT	2640
Db	2738	GATGAGACTTGCAGAAACAGGGGAAAAATTAGTAAACAGATATGTCACTGTAGACTT	2797
Qy	2641	GCTGCTGGGAAGTCTATTGTGTAGAGTGGCACAGAACCTGTATTACCATGAATCATCA	2700
Db	2798	GCTGCTGGGAAGTCTATTGTGTAGAGTGGCACAGAACCTGTATTACCATGAATCATCA	2857
Qy	2701	TTAGAACAAATATACGCTATGTGCATTTGCTATTCACACGATGAAATGCTATACAGTAAAGCAA	2760
Db	2858	TTAGAACAAATATACGCTATGTGCATTTGCTATTCACACGATGAAATGCTATACAGTAAAGCAA	2917
Qy	2761	GTGTTTGGCCAGAAACTTCACAAAGGCTTTCGCCGTTTACGGCTTCCACTGAGTATATG	2820
Db	2918	GTGTTTGGCCAGAAACTTCACAAAGGCTTTCGCCGTTTACGGCTTCCACTGAGTATATG	2977
Qy	2821	GCAATCTGTGCCCTTTGTGCAAAAAGATCTGTAAAGAGAGAAGACTCATGTAGGCAA	2880
Db	2978	GCAATCTGTGCCCTTTGTGCAAAAAGATCTGTAAAGAGAGAAGACTCATGTAGGCAA	3037

QY	2881	TGTTGGTGAATAATATTAATGTAGAGCGGGAGATCTGGAAGCAGCATGACGCTTTAGT	2940
Db	3038	TGTTTGTGAAAAAATATTAATGTAGAGCGGGAGATCTGGAAGCAGCATGACGCTTTAGT	3097
QY	2941	GAAAAATATGTGCTCTCTACCGAGATATGTGTCCATATACAAATCACCCTTTGGCA	3000
Db	3098	GAAAAATATGTGCTCTCTCTACCGAGATATGTGTCCATATACAAATCACCCTTTGGCA	3157
QY	3001	CATGACCCGAGATTATGTCAAGATACAGATATTGACAACTTAAAGATTTAAAGAAATGT	3066
Db	3158	CATGACCCGAGATTATGTCAAGATACAGATATTGACAACTTAAAGATTTAAAGAAATGT	3217
QY	3061	CTTGGTTGTTGTTCTGGAATATATATATGGCTTAAAAATGAAATTAACAGTCACGCTTTTATC	3120
Db	3218	CTTGGTTGTTGTTCTGGAATATATATATGGCTTAAAAATGAAATTAACAGTCACGCTTTTATC	3277
QY	3121	AGAAAGATGTAGAAATATTAAACCAACAAAGATGGCCCAAGGACCGAGATGACAAAA	3180
Db	3278	AGAAAGATGTAGAAATATTAAACCAACAAAGATGGCCCAAGGACCGAGATGACAAAA	3337
QY	3181	ATGAATGAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAT	3240
Db	3338	ATGAATGAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAT	3397
QY	3241	ACTAATATCAGTTTGGATCTCTTAAAGACCGGGATACACAGCGCTGTTCTACATCAA	3300
Db	3398	ACTAATATCAGTTTGGATCTCTTAAAGACCGGGATACACAGCGCTGTTCTACATCAA	3457
QY	3301	CCTGACAGAAATTTTCAGTACACCAAAAAATTATCTGCTCTGTGAATGAATTCATTTTC	3366
Db	3458	CCTGACAGAAATTTTCAGTACACCAAAAAATTATCTGCTCTGTGAATGAATTCATTTTC	3517
QY	3361	ACTCTGGAAAAACCTAAACCAACCATGTCTAGAGAGCTGTAAACCCACTTTCATCA	3422
Db	3518	ACTCTGGAAAAACCTAAACCAACCATGTCTAGAGAGCTGTAAACCCACTTTCATCA	3577
QY	3421	GCAGGCAAGCATCTCAGACCAAAATCATCAGAAATGGAAGCTGTAAAGCATGCAAGCAGC	3486
Db	3578	GCAGGCAAGCATCTCAGACCAAAATCATCAGAAATGGAAGCTGTAAAGCATGCAAGCAGC	3633
QY	3481	AGCTCAATCCAAAGCTCTCTCTGTGAAGATTAAGGGGAGGCTTGATAGTTCGAAATGCAT	3540
Db	3638	AGCTCAATCCAAAGCTCTCTCTGTGAAGATTAAGGGGAGGCTTGATAGTTCGAAATGCAT	3697
QY	3541	CACAGTGAATAATGAAGATTACACAATGCTCTCACCTTTGCCGGGAAAAAAGTGCACAG	3600
Db	3698	CACAGTGAATAATGAAGATTACACAATGCTCTCACCTTTGCCGGGAAAAAAGTGCACAG	3757
QY	3601	AGAGACGACTGTGATCTTTGAAGGCTGTGAATTTGGAAAGCCTTAGAGCAGGAAAAAAACG	3666
Db	3758	AGAGACGACTGTGATCTTTGAAGGCTGTGAATTTGGAAAGCCTTAGAGCAGGAAAAAAACG	3817
QY	3661	CCCGTCACAGAAACAGGAGAGAAATTAGTATGATGATGACTGAATATGTTGTCACAGAA	3720
Db	3818	CCCGTCACAGAAACAGGAGAGAAATTAGTATGATGATGACTGAATATGTTGTCACAGAA	3877
QY	3721	CAGAAACCTTAAAGGACGTCAGCGAAGTGGGAAAAAGAGGCCATACGGCTTCAGAAATCTGAT	3780
Db	3878	CAGAAACCTTAAAGGACGTCAGCGAAGTGGGAAAAAGAGGCCATACGGCTTCAGAAATCTGAT	3937
QY	3781	GAACAGCAGTGGGCTGAGAGAAAAAGGCTCAAAAGAAATATATTNAGAAATGAAGATGAA	3840
Db	3938	GAACAGCAGTGGGCTGAGAGAAAAAGGCTCAAAAGAAATATATTNAGAAATGAAGATGAA	3997
QY	3841	CAGAATATCTCGCCCAAAAAAGGTTAAAGAGGCGACCAACCAAAACCTCTTGATGAGAGT	3900
Db	3998	CAGAATATCTCGCCCAAAAAAGGTTAAAGAGGCGACCAACCAAAACCTCTTGATGAGAGT	4057
QY	3901	ACACCAAAAGAGAGCCCAACAATGAAAACTTTAAAAAAGGAGACCAAAAAAAATCTTGA	3966
Db	4058	ACACCAAAAGAGAGCCCAACAATGAAAACTTTAAAAAAGGAGACCAAAAAAAATCTTGA	4117

Db 833 CATGATTGCCAATTTTTCATCAATCAAGTCGCTGGGAAATCATCAGTAGATT 892
Qy 752 TGTGAGACATGCTCTTGTACTTAATTTTGAGCTCTACAAATATATGATGATCTTGTCTC 811
Db 893 TGTGAACATGATTTGATCTGATTCAGAACTTTTGTATATATCTCTCTATATAT 952
Qy 812 TCTCTGTTTATCCAGCTTGATTTAAATTAAGACAGATGATTAAGAGCGCTTAC 871
Db 953 TATCGGTATGCGACAGCTTGATTTCAAACTTAAGCAATGATGAGAGAGCATTAG 1012
Qy 872 AAGTTGTAACCTACTGCAAAAATTTGGGGCAAGAGATTCAGATTTGGCTCTCAAA 931
Db 1013 CTGTGTTCGACTTACCTAAATTTGCTGCTCAAAAGATTTCTAATTTGGCAACAG 1072
Qy 932 ACAAGCCATTGGCAGTCTACTTGGCAGGTTTAATGATATCCATGATACCAATCCCC 991
Db 1073 ATCGTCCTTTGGCAATGTTTCTTTGGACGATTTAATGATATTCATGTCCTGTGAGAT 1132
Qy 992 TGGAAATGTGAAATTTGCTAGCCATTTGCTCATGACCAATCCGATTTAGCAAAAGCT 1051
Db 1133 TAGAAAGTGTGAAATTTGCCAGTCAATGTTTAAATGAAATCACCCAGATTTAGCGAAGATC 1192
Qy 1052 TAACAGATATCTTAAAGTGAAGTGAATGACGACGAGGAGGAGGATTTAGACATGATGTA 1111
Db 1193 TCACGAAATATTTAAGGTATGATACATGATCCAGAAAGCTAATTCGTCATGATGCA 1252
Qy 1112 TTGTCTCAATAGTTACAGCTGCTAAAGAGATATCTTCTGTGTCATGATCACTACTTA 1171
Db 1253 TTGTTACTATATATAACGCTGCGAAGAGGACCTGGCTTAGTAATATGATTCAGCTGCTTG 1312
Qy 1172 ATTTTGTAGAGAGAGAACTTTAGCAAAACGATGAGAGTACGCAAGAGCCATGATGG 1231
Db 1313 GCTTGTGTAAGGAAAGACACGATGAACGCTGGCGAGTAAAGAGAGGATATGATGG 1372
Qy 1232 GACTTGGCCCAATTTATAGAAATATGCTTTACAGTCAAGCAGCTGGAAGATGGTGA 1291
Db 1373 GTCTGCTCAGCTTTATAGAAATATCTCTTCATGATGAGAGAGAGAGAGTGCAG 1432
Qy 1292 AACAGATAGCATGATCAAGAAACAAATTCATATATATATCAAAATATGATTTGATG 1351
Db 1433 AGAAAGTACGCTGGATTAAGAGCAAACTTCGCAATTTATATATCAAGAACATTTAGAG 1492
Qy 1352 ATCGCTACTTGTTAACGAGATCTTGTCTCAATACATGCTTCTCAATTTAGAACTA 1411
Db 1493 ACAAACTGTGTGTAAGAAATCTTGTCTCAGTATCTTGTCTCCCAACACCTGGAAAG 1552
Qy 1412 CAGAGGATGAATGCTTATATATCTTGTATGCGACACTGGAATTAATGCTGGAAG 1471
Db 1553 AAGAGAGATGAAATGCTTATATATCTTATATGCTAGTTGGATCCAAATGCTGTAAG 1612
Qy 1472 CATGGAATGAATGTGAATGTCAAAATCTGCTCCGACATCAAGTAAAGATTTGCTTG 1531
Db 1613 CTCTCAAGAAATGGAAAGTGTGCAACATGCTTCGAGGCCATTTAGCGCAATTTGG 1672
Qy 1532 ACTTGATTAAGCAACCCAAACAGATGCCAGTGTAAAGCCATTTTCCAAAGATGATG 1591
Db 1673 ATTTGACAAACAGCTATCATCAGAGGCTTAACGTCTTCCATGCTTTGGAAGAGATG 1732
Qy 1592 TTATTAACAAGAAATTTACCTGATCCTGTAAGGCTCAGAGATTTCAATGAAGAAATCAC 1651
Db 1733 CCATAGCAAGAAATTTGCTCTGACCCCGGAAAGCAAGATTTTGTGAAGAAATTTAAC 1792
Qy 1652 AGGTGTTAGAAATGATGAGAAATTAAGAAAGCAGTTAGAGTACTTGTAGTCCACAT 1711
Db 1793 AGTTTCTCGCGAGATGAGAAACTTTCGCTCAATGAGATTTATTAATTTAGCCAACT 1852
Qy 1712 GCTCTGTGAAAGCAGCTGAAGGTGTGTGCTGAAATTACTTAAGAGTTGGCAACCCCA 1771
Db 1853 GTTCTTGCAACCAAGCAATATTTGTGTGAGAGAAATGCCCCGGAACCTTCGCAATTCCTA 1912
Qy 1772 AACAGCCTACAAATCTTCTCGAAATGATCAAGTTCTCTGAGAGAGATGACGCTG 1831
Db 1913 AGCAACCAAAATCTTTCTAGAGATGTCAAAATTTCTGTGGAAGAAATCGCACCTG 1972

Qy 1832 TGCACATAGATACCGAATCTATACAGTGTCTTATTAACAAAGTGAACAAATCAATGATG 1891
Db 1973 TGCACATGATTTACAAAGCCAAATAGTGCACATGTGAATTAATGAAATGCAATGAGG 2032
Qy 1892 GACACGAGATGATGAAGTGAAGGTGTTCCACTGATCAACGCCATCAGAGCAGTCTTG 1951
Db 2033 GGACAGCAGATGATGAAGGAGGAGGTGTAAGTCCAGATACAGCTATCCGTTCCAGACTTG 2092
Qy 1952 AACCTGTAAGTACTGATTCATTCACATCCCATCTCATCTTCAATTCCTGCTGAACATTTG 2011
Db 2093 AACTCTTAAGTCTGTCTTTTACACATCTCACTCGTCTGCTCCTGTGAGAGATATG 2152
Qy 2012 AATCATTTCTGCTGTGTGAATAATGATGATGAATAAAATAGACAGAGCTGACTACAA 2071
Db 2153 AGTCTTGTATAGAGCCTAAGAATGAGAGATGACAAAGGTATACAGAGCTGATTTCAA 2212
Qy 2072 TTTTCAAAAACAGAGAAATTTGAAGAGATTTTCCACATTCAGATCAGATCGCTTGC 2131
Db 2213 TTTTGAAGAAATACAGCTCAAAATTAAGAAACAGACCTTCCAGATGATGATGACCTTAA 2272
Qy 2132 TTCCGTTTACATCAAAATCTAAGAAAGGACCCCGCCGACAGCCAAATATGCAATTC 2191
Db 2273 TTCCATTTTACATCAAAAGCAAGAGAGGATCTCACACACCAAGCAAAAGCTGTGC 2332
Qy 2192 ATTTATTCATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTC 2251
Db 2333 ACTGATATACAGCCATATTTCAAAATTAAGAAAGTCCAGCTTGCACAGATTTTGGAGCAC 2392
Qy 2252 TGCATTAAGACCTACATCCAAAGCAACCTGGAACATCTATTAACCATTTGGTACTATG 2311
Db 2393 TCAGTAGAGAGTGTGATCTGATGTGCGAAGCAAACTTAATCTCATTTAGTTCTATGG 2452
Qy 2312 GTCATATTTGCTCTGCTGACCTGATCAATTTGCTGCTCTGGAATCTGGTAGCTA 2371
Db 2453 GCCACATTTCTATGTAGCACAGATCAAGTTGCTTCCCAATGAATCTGTAGTAGCA 2512
Qy 2372 CTTTCATTTGTGAAGATCTTCTCATGAATGATGCTGCTTCCAGAGGAAAGCAACATTAAC 2431
Db 2513 ATTTTATTTGTAAGATCTGCTAATGATGACAGTCAACAGGTGAAGAAATGAGAAAC 2572
Qy 2432 TTTGGGTTCCAGATTAAGATATCTCTGAGACATGCTCAAAATTCAGCTATTAATA 2491
Db 2573 TGTGTTCTCCAGATTAAGAGTTTCCCTGAAGTCTAGCAAAAGGTGAAGCAATTTAAC 2632
Qy 2492 TGATGTTCCAGATGCTTGTGGAATGAATAATACAGATTAATCAGAGATCTTACT 2551
Db 2633 TTCTGTGAAGTGGCTGTGGGTATGAAAAACACCATCTTAATCTGCCAATTTAACCC 2692
Qy 2552 TAAGATTGCTAACCAATATTTGCATAGTGTGAGAGCTTGACAGAGAGGGGAAATTA 2611
Db 2693 TTCCGTTATTTACAGAGATGTTGGTTAGTGAGAGGTGACCTGACAGAGAAAGAGATCA 2752
Qy 2612 GTAACAGATATGTCAGCTGTGAGACTTGTGCTGGAGAGTCTATTTGAAAGCTGGAC 2671
Db 2753 GTAAATGATATGCTGCTGCTGATGAGTGTGCTGAGTGTGCTCAATTAAGTGTGCTC 2812
Qy 2672 AAGAACCTGTATACATGAATATCATCATTAAGAACATATACAGTATGATGCAATAGCTA 2731
Db 2813 AGGAACCTGTATACATGAATATTTACCCAGAAACATTTTACAGTCTGTGCTACCTGTGA 2872
Qy 2732 CCCGTTACGGCTTCCACTTGTAGTATATGGAATCTGTGCCCTTTGTGCAAAAGATCTG 2851
Db 2933 TGAAGTACTACTCCCATTTGAGATATATGAGGATCTTTGCTGTGTGCCAAAGTCTCTG 2992
Qy 2852 TAAAGAGAGAAAGACCTATGCTAGGCAATGTTGGTGAATAATTAATGAAGCGGG 2911
Db 2993 TGAAGAGAGAAAGACACAGCAGCAATTTTACTGAATAATATTCAGTATACGAGGG 3052

Db	414	TGTTCTGGCATGTTTGGAAAACACTGATGACATAGCAAAAGATTTGCTTGACCCCGGAAA	473
OY	1624	GCTCAGATTTTCATGAGAATAATTCACACAGGTGTTAGAAATGATGAGAAAAATTAAGAAAG	1683
Db	474	GCACAGAGATTTTGTGAMAGAAATTTAAACAGGTTCTCGCGAGATGATGAGAACTTGCGTCT	533
OY	1684	CAGTTAGAGAGTACTGTTAGTCGCAACATGCTCCCTGCAAGCAGCGCTGGAAGTTGTGTGGT	1744
Db	534	CAGTTGGAGTTATTATTATTAATGACCAACCTGTTCTTCCAAACAAGCAGATTAATTTGTTTGGA	593
OY	1744	GAATTAACTAAGAAGTTGGGCAACCCCAACACAGCCTTACAAATCTTCTCTGGAAATGATC	1803
Db	594	GAATATGCCCCGAAACTTGCAAACTCTTAAGCAACCAACAATCCCTTTCTAGAGATGGTC	653
OY	1804	AAGTTTCTCTTGGAGAGATAGCACCTGTGCATATGATACCGAATCTATCAGTGCCTT	1866
Db	654	AAATTTCTGTTGGAAAGAAATGCGACCTGTGCATATGATTCAGAAAGCCATTAAGTGCCTA	713
OY	1864	ATTAAACAGTGAACAAATCAATAGATGGAACAGAGATGATGAAGATGAGGCTGTGCA	1922
Db	714	GTGAATTTGATGAATAAATCAATAGAGGGGACAGCATGATGAAGAGAGGTTAAGT	773
OY	1924	ACTGATCAAGCCATCAGACAGGCTTGTAACCTGTTAAGATCTCTCATTTACACATCC	1983
Db	774	CCAGATTAACGGTATCCGTTGACAGACTTAACCTTCTTAAGGTTCTGTCTTTAACACATCT	833
OY	1984	ATTCATATTCTGCTGTAACAAATTTGAATCATATACGGGCTGTCTGTAATAATGGATAT	2043
Db	834	ACCTGTTCCACTCTGCGAGACATATATAGTCTTGTTCACGTGCTTAAGATTTGAGAT	893
OY	2044	GAATAAGTAGAGAAAGCTGCACATACAAATTTTCAAAAACACAGAGAACAAATTTGAAG	2103
Db	894	GACAAGTAGAGAAAGCTGATTCACAAATTTTGAATAATACAGGTCACAAATATAGAAC	953
OY	2104	GATTTTCCACATCATGATTCAGCCCTGCTGCTGTTTACATTCACAAATCTTAAAAAG	2166
Db	954	GACCTTCCCAAGATATGATGATGACCTTTAATTCCTCATTTTACATCAAAAAGCAAGGGGT	1013
OY	2164	CCCCCCCCGACCAAAATATGCCATTTATGATATCCATGGGATATTTTCTAGTAAAG	2223
Db	1014	ACTCCACACCAAGCAAAACAGGCTGTGCACGTATACAGGCCATTTTCACAATTAAGAA	1073
OY	2224	ACCAGTTTGCACAGATATTGAGCCTTGCATAGAAGCCTTAGATCCAAACCTGGAA	2288
Db	1074	GTCACGCTTGCACAGATTTTGTGACCCACTCAGTAGGATCTGAAAGCTGATGTGCAACAA	1133
OY	2284	CATCTCATACACACTGTTACTATTGGTATATTTGCTCTCTGCACTGATCAATTT	2343
Db	1134	CAATTTAATCCTCATATGTTTCAATTTGGGCCACATTTCTATGTTAGCCACAGATCAGTTT	1199
OY	2344	GCTGCTCCTTGGAAATCTTGGGTAGCTATCTTCATTTGTGAAGATCTTTCATGATGAT	2403
Db	1194	GCTTCCCAAGAAATCTGTAGTAGCAAAATTTTATGTGAAGATCTCTTAATGATGATC	1253
OY	2404	CGGCTTCCAGGGAAGAAAGCAACATAAATCTTGGGTCCAGATGGAAGATATCTCCCTAG	2466
Db	1254	AGGTCACAGGTGAAGAAATGAAAGAACTGTGGTCCAGATGAAGAGGTTTCCCTGAA	1313
OY	2464	ACAAATGGTCAAAATTCACAGCTATTAAATGATGTTGGTGCATGCTACTTGGATGAAAAAT	2523
Db	1314	GTACTAGCAAGTAGACAGCAATTAATCTTGGTAAGGTGGTGTGGTATGAAAAC	1373
OY	2524	AATACAGATAATCAGAACTTCTACATTAAGATTTGCTAACACAAATATTCATATGTGAT	2588
Db	1374	AACCAAGTCTAATATGCCCCAATTCACACCTTGTTGATTAATACAGGATTTGGTATGGAG	1433
OY	2584	GGACACTTGACAGAAACAGGGAATAATTTGTAACACAGATATGTCACGCTGACAGCTTCT	2643
Db	1434	GGTGACCTGACAGCAAAAGAGATCAGTAAATCTGATATGTCCTGCTTCCGATTAAGCT	1493
OY	2644	GCTGGAGTCTATTGTGAAGCTGGCACAAAGACCTTGTATCCATGGAATCATCATTA	2703

Db	1494	SGTGGTAGTGCATATATGACCTTGCTCAGGACCTTGTATCCATGAAATATTATACCCA	1553
Oy	2704	GAACAATATCAGCTATGTGCACTTACGATCAACGATGAATGCTATCAAGTAAGCAAGTG	2763
Db	1554	GAACAGTTTACAGCTCTGTGCACCTGTTATTTAATGATGAGTGTACCAAGTAAGCAGATA	1613
Oy	2764	TTTGGCCGAAGAACTTCACAAGGCGCTTCCGCTTTACGGCTTCCACTTGAAATATATGGCA	2823
Db	1614	TTTGGCTGGAAGCTGCATAGGCACTTGTGAAGTTACTGCTCCCATTTGGAGTATATGGCG	1673
Oy	2824	ATCTTGCCCTTTTGTGCAAAAGATCCCTTAAAGGAGAGAGAGCGTCATGTGAGCCAAATGT	2883
Db	1674	ATCTTGCTGTGTGTCGCAAAAGATCCTGTGAAGGAGAGAGAGACACAGCAGCAATATGT	1733
Oy	2884	TTTGGTGAAGAAATATTAATATGTAAAGCGGAGATATCTGAAGCAGCATGCGACTGTTATGCA	2943
Db	1734	TTTACGAAAAAATATCAGTATACGCGAGGAATATCAATTAAAGAAATCCATATGGCTATCAG	1793
Oy	2944	AAATATTGTGCTCTCTACCAAGATGTTGTTCATATCAATTAATCACTTTGGCCAT	3003
Db	1794	AAATATTATATCAGTGTGCTGCGTAATATATAGTTCCATATCAGATTTACACTGCTGACCAT	1853
Oy	3004	GACCCAGATTTATGTCAAAAGTACAGGATATTGAACAATTAAGATGTTAAAGATGTCTT	3063
Db	1854	GATCCAGATTTTACCAAGATACCAAGATGTTGATCAGTTCCTGTATATCAAAAGATGCTTA	1913
Oy	3064	TGTTGTGTTGGAAATTTAATGAGCTTAAATGAAATTAACATACATCAGTCAGCTTTATACGA	3123
Db	1914	TGTTTCATGCTTTGAAGTTTATATGACAAAGATGAAATCAATATGCCATGCTTTATGAG	1973
Oy	3124	AAGATGTTAGAAATATTTAAACAAACAAAGATGCCCAAGGACAGATGATGCAAAAATG	3183
Db	1974	AAGATGGCAGAGAAACATCAAGTTAAACAGAGATGCCAGTCCAGATGAATCCAAAGCA	2033
Oy	3184	AATGAAAACTGTACACTGTGTGTGATGTGCCATGAATATCATCATGTCAAAAGTACT	3243
Db	2034	AATGAAAACTGTATACAGTATGTGATGTGCTCTCTGTATTATTAATATGATAAAGTGT	2093
Oy	3244	ACATACAGTTTGGATCTCTCTTAAAGACCCGGTACTTACAGGCTCGCTTCTTCTACCACT	3303
Db	2094	TTGTGCATATGCAGATTCACCAAGGAGGCCAGTCCCTCCCAATGAATTTTTTACACAACT	2153
Oy	3304	GACAAGAATTTACATGACACCAAAAATATATGCTCTCCTCAATGAATCATATTTTACT	3363
Db	2154	GAAAAAGGACTCTGTGAATGAATTAAGATTATTTTCAGAAAGACAAAGATACCTCTGTTA	2213
Oy	3364	CCTGGAACCTTAAACCAACCAATGTTCTTAGAGCTGTATTAACAGCCACTTTCATCGCA	3423
Db	2214	ACAGGAAGCCCAAGGCCGTGTGAGTACTATAGGTGCAGTAAATTAAGCCTTTATACGACAG	2273
Oy	3424	GCGAAGCATCTCAGACCAAAATCATACGAATGGAACCTGTAAAGCAATGCAAGCAGAGC	3483
Db	2274	GGAAGGAAGACCTTATGTTAGAAGCACTGGCACTGAGACTGAAACCAATATTAAATGTAAAT	2333
Oy	3484	TCAATTCOAACCTCTCCGGAAGAAATAAAGGGAGGCGTGTATTTCTGGAATGATACAC	3543
Db	2334	TCAGAGCTGAACCTTTCACCGGAAATCGATCAAGGAGACAGAGATTCAGAGGACGACGAA	2393
Oy	3544	AGTGAATAATGAAGA	3557
Db	2394	ACTGGAGTTAGTGA	2407
RESULT 8			
AAAA7423			
ID	AAAA7423 standard; DNA: 3957 BP.		
XX	AAAA7423;		
AC			
XX	20-OCT-2000 (first entry)		
DT			
XX	Sequence encoding human neuron-associated protein.		
DE			

KW	Neuron associated protein; NEUP; neurological disorder; epilepsy;
KW	ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW	Alzheimer's disease; Pick's disease; Huntington's disease;
KW	dementia; Parkinson's disease; demyelinating disease; meningitis;
KW	peripheral palsy; muscular dystrophy; central nervous system; CNS;
KW	peripheral nervous system; PNS; myopathy; schizophrenia;
KW	actinic keratosis; arteriosclerosis; atherosclerosis; burns;
KW	clonus; hepatitis; mixed connective tissue disease; MCTD;
KW	myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW	autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW	AIDS; Addison's disease; adult respiratory distress syndrome;
KW	allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW	Werner syndrome; trauma; human; ds.
XX	Homo sapiens.
OS	
XX	
XX	
FH	Key
FT	Location/Qualifiers
FT	140..2704
FT	/tag=a
FT	/product= Neuron associated protein
XX	
PN	MO200034477-A2.
XX	
PD	15-JUN-2000.
XX	
XX	10-DEC-1999; 99WO-US30408.
PF	XX
XX	
PR	11-DEC-1998; 98US-0210083.
PR	01-DEC-1998; 98US-9123456.
PR	09-FEB-1999; 99US-0119365.
PR	16-MAR-1999; 99US-0124687.
XX	
PA	(INCY-) INCYTE PHARM INC.
XX	
PI	Tsang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J.
PI	Lu DM, Azimzal Y;
XX	
XX	WPI; 2000-423423/36.
DR	P-PDB: AAB01382.
DR	
XX	
PT	New human neuron-associated proteins and polynucleotides encoding them
PT	useful for diagnosing, treatment and prevention of cell proliferative
PT	disorders including cancer, neuronal and neurological disorders
XX	
PS	Claim 9; Page 130-131; 145pp; English.
XX	
CC	Human neuron-associated proteins (NEUP) can be used for for
CC	treating or preventing a disorder associated with decreased
CC	expression or activity of NEUP. Antagonists of NEUP are useful for
CC	treating or preventing disorder associated with increased expression
CC	or activity of NEUP. NEUP or their fragments or derivatives are
CC	useful for treating neurological disorder such as epilepsy, ischemic
CC	cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC	disease, Pick's disease, Huntington's disease, dementia and
CC	Parkinson's disease. NEUPs are also useful for treating other
CC	demyelinating diseases, bacterial and viral meningitis, prion
CC	diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC	metabolic diseases of the nervous system, neurofibromatosis, other
CC	developmental disorders of the central nervous system, cerebral
CC	palsy, neuromuscular disorders, autonomic nervous system disorders,
CC	cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC	other neuromuscular disorders, peripheral nervous system disorders,
CC	inherited, metabolic, endocrine, and toxic myopathies, mental
CC	disorders including mood, anxiety and schizophrenic disorders, a cell
CC	proliferative disorder such as actinic keratosis, arteriosclerosis,
CC	atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC	tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC	haemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC	bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC	disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC	disease, adult respiratory distress syndrome, allergies, ankylosing
CC	spondylitis, amyloidosis, anemia,asthma, Werner syndrome,
CC	

CC	complications of cancer, hemodialysis, and extracorporeal circulation,
CC	viral, bacterial, fungal, parasitic, protozoal, and helminthic
CC	infections, and trauma. This sequence was given the Inctye ID no.
CC	2888437CBI.
xx	
50	Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;
	Query Match 26.7%; Score 1112.8; DB 21; Length 3957;
	Best Local Similarity 68.1%; Pred. No. 3.5e-264;
	Matches 1549; Conservative 0; Mismatches 727; Indels 0; Gaps 0;
QY	1282 GATGCTGAAAACAGATAGCATGATCAAAAGCAAAATTCCTACATATATATATCAAAAT 1344
DB	2 GAAGCTGGAGAGAAAGTACGCTGGATTAAGGCAAACTTCGCAATTTATATATACGAC 61
QY	1342 AGATTTAGATGATCGACTACTGTTGTAAGGAGATCTTGGTCAATACATGATGTTCCACAAAT 1400
DB	62 AGCATTTGACGACAAACCTGTTGTAGAGAAATCTTTGCTCAGTATCTTGTCCCAACAC 121
QY	1402 TTAGAAATACAGAAAGGATGAAATGCTTATATATCTTGTATGCCACACTGGATTTAAAT 1465
DB	122 CTGGAACAGAAAGAGAGAAATGCTTATATCTTATATCTTATGATTTGGATTCCAAT 181
QY	1462 GCTGTGAAGACATTCATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAG 1522
DB	182 GCTGTAAAGGCTCTCAACGAAATGTGGAAGTGTCAAGAACATGCTTCGGATCCATCTAGCC 241
QY	1522 GATTTGCTTGATTCATTTAGCAACCCAAAACAGATGTCAGTGTCAAGGCCATATTCCA 1589
DB	242 GACATTTGATTTTGCACACACAGCTCAATCAGAGGCTTAACCTGTTCCCATGTTTGA 301
QY	1582 AAAGTGATGTTTATTCACAGAAATTTACCTGATCCCTGTAAGGCTCAGAGATTTTCATGAG 1644
DB	302 AAATGATGATTCATTCACAGAAATTTTGGCTGACCCCGGAAAGACAAAGATTTTGTGA 361
QY	1642 AAATTCACACAGGTGTGGAAGATGATGAGAAATTAAGAAAGCACTTGAAGTACTTGT 1700
DB	362 AAATTTAACACAGGTCTCGCGAGATGATGAGAAACTTCGCTCACTGGAGATTTAAT 421
QY	1702 AGTCCAAATGCTCTCGAAGGAGCTGAAGGTTGTGCGGAATTAACATTAAGAAATG 1765
DB	422 AGCCCAACCTGTTCTTGCAAAACACCAATATTTGTGTGAGAAATAGCCCGGAACCTT 481
QY	1762 GGCAACCCCAACAGCTACAAATCTTCTCGTGAAGATCAAGTTCCTCTTGGAGAG 1821
DB	482 GCAATTCCTTAAGCAACCAAAATCTTTCTAGAGATGTCAAAATTTCTGTGGAAGA 541
QY	1822 ATAGCACCTGTGCACATAGATACGAATCTATCAGTGTCTTATTTAAACAAGTGAACA 1881
DB	542 ATGCACTGTGCACATGATTTACAGAACCCATTAAGTGCATGTAATGTAATG 601
QY	1882 TCAATAGATGCAACAGCAGATGATGAATGAGGTTGCCAATGATCAAGCCATCA 1944
DB	602 TCAATAGAGGGGACAGCATGATGTAAGAGGGGTGTAAGTCCAGATACAGCTTCCGT 661
QY	1942 GCAGGCTTTGAACCTTAAGTACTTCATTTACATCCATCTCAATTCATCTCTGCT 2004
DB	662 TCAGGACTTGAACCTTTAAGGTTGTGTCTTTTACACATCTTACTCTGTCACCTGTCA 721
QY	2002 GAAACATTTGAATCAATTAAGTGTGCTTGAAGAAATGATGATGAAGAAAGTATGACAGAGCT 2065
DB	722 GAGACATTTGATGCTCTTGTACAGTGCTTAAGAAATGAGAGATGACAAAGTATGACAGAGCT 781
QY	2062 GCACATCAAAATTTTCAAAAACACAGAGCAAAATTTGAAGAGATTTTCCACATCA 2122
DB	782 GCTATTTCAAAATTTTGAAGAAATACAGGTCACAAATTAAGAACAGATCTTCCAGATTA 841
QY	2122 TCAGGCTTGTCTCGTTTATATCAACAAATCTAATAAAGAACCCCGCTCAACGCAAA 2181
DB	842 TCGACCTTAATTTCCATTTTACATCAAAAGCAAAAGAGGGGTACTCCACACCAACAAA 901
QY	2182 TATGCAATTCATTTATCCATGCAATATTTTCTAGTAAAGAGACCACTTTGCACAGATA 2241

Db 902 CAGGCTGTGCATGTATACAGCCATATTCACAATAAAGAGTCCAGCTTGACAGATT 961
 QY 2242 TTGAGCCTCTGCATTAAGAGCCTTAGATTCAGCAACCACTGAGACATTCATACACATG 2301
 Db 962 TTTGAGCCACTCATAGAGAGTCTGATTCGTATGTCACAGAACCACTTATATACATCA 1021
 QY 2302 GTTACTATGTCATATTTGCTCTCCCTTGACACCTGATCAATTTGCTGCTTGGAAATCT 2361
 Db 1022 GTTTCATTTGGGCGCACATTTCTATGTGTAGCACAGATAGTTGGCTCCCAATGAATCT 1081
 QY 2362 TGGGTACTTCTTCTGTAAGATCTTCTCATGATGATGCGCTCCGAGGAAAAG 2421
 Db 1082 GTACTAGCAATTTTATTTGTAAGATCTGCTAATGATGACAGGTCAACAGGGTGAAG 1141
 QY 2422 ACAACTAAACTTTGGGTCCAGATGAGAAAGTATCTCTGAGACATGTCGCAAAATTCAG 2481
 Db 1142 AATGGAAGACTGTGGTCTCCAGATGAGAGAGTTCCCTGGAAGTACTGCAAGAGTACAG 1201
 QY 2482 GCTATTAAATGATGCTTCCATGGCTACTTGGAAATTAATCACAGTAAATTCAG 2541
 Db 1202 GCAATTTAACTCTGTGTAAGGTGGCTGTGGGTATGAAAAACAACAGTCTAAATCTGCC 1261
 QY 2542 ACTTCTACCTTAAGATGCTTAACAACAATATTCATGATGATGGAGACTTGACAGAACAG 2601
 Db 1262 AATTCAACCTTCCGTTATTTATCAACGATGTGGTACTGAGGCTGACCTGACAGAGCAA 1321
 QY 2602 GGGAAATTTAGTAAACAGATATGTCACGTCTGAGACTTGGCTGGAGTGTCTATTGTG 2661
 Db 1322 AAGAGATACAGTATGATGATGTCGCTGGGATAGCTGGGATAGCCATATAG 1381
 QY 2662 AAGTGGCACAAGAACCTGTTCACATGAATTCATCATTTAGAACATATACAGTATGT 2721
 Db 1382 AAGCTTCTCGGAACCTGTGTACCATGAATTTATCCCGACAGAACATTTACACTCTGT 1441
 QY 2722 GCATTACTATCAACGATGATGATGCTATCAATGATGAGACAGATGTTGCCCAAACTTCA 2781
 Db 1442 GCATTTGTTATTTATGATGATGATGTTACCAAGTAAAGCAGATATTTGCTCAAGAGTGT 1501
 QY 2782 AAAGGCTTTCCCTTTACGAGCTTCCACTTGTAGTATGAGCAATCTGTGCCCTTTGTGCA 2841
 Db 1502 AAGGCACTTGTGAAGTACTCTCCCATTTGGAGTATATGGGATATTTGGCTTGTGCC 1561
 QY 2842 AAAGATCCTGTAAGGAGAGAGAGAGAGCTCATGAGCAATTTTGGTGAATAATATTAAT 2901
 Db 1562 AAAGATCCTGTAAGGAG 1621
 QY 2902 GTAGGCGGAGATCTGATGAG 2961
 Db 1622 ATAGCAGAGAGATATCATTAACAGAGATGCTATGCTACTGAGAAATTTATCACTGTG 1681
 QY 2962 CCAGAGTATGTTGCTCATATATCAATTCACCTTTTGGCAGATGACCCAGATTATGTCAA 3021
 Db 1682 CCTGAATATGTAGTCCATATCATATTCACCTGTAGCCCATGTATCCAGATTTTACAAGA 1741
 QY 3022 GTACAGGATTTGAACAACATTAAGATGTTAAAGATCTTTGGTTGTCTGGAATA 3081
 Db 1742 TCACAAAGTGTGTATCAGCTTCTGTATATCAAGAGTCCATGATGCTATCTGGAATTT 1801
 QY 3082 TTAATGCTAAATTAAGAAATTAACAGTACAGCTTTTATCAAAAGATGTTGAAATATTT 3141
 Db 1802 TTAATGACAAAGATGAAATAATATGACATGCTCTTTATGAAAGAGATGCGCAGAGACATTC 1861
 QY 3142 AAACAAACAAAGATGCGCAAGAGACAGATGATGCAAAATTAATGAATAAACTGTACACT 3201
 Db 1862 AAGTTAACCAGATGCGCAAGTCTGCAATGATGATGATGATGATGATGATGATGATGAT 1921
 QY 3202 GTGTGTATGTTGCCATGATATCATATCATGTAAGAGTACTACATACAGTTTGGAAATCT 3261
 Db 1922 GTATGTGATGTGGCTCTGTGTATATAATAGTAAAGATGCTTGTGCAATGACAGATTTCA 1981
 QY 3262 CCTAAGACCGGATGATGACAGCTGCTTCTTCACTCAACCTGACAAATAATTTGATGATAC 3321
 Db 1982 CCAAGGACCGAGTCTCCCAATGAAATTTTATACACAACTGAAAAAGACTTCTGTATAC 2041

QY 3322 ACCAAATATATCTGCTCTCGAATGAATATTTTCACTCTCGGAAAACCTTAACA 3381
 Db 2042 GATAGAGTTATATTATGAGAAGACAGAGAGTACTTCTGTTAACAGAGAAACCAACCT 2101
 QY 3382 ACCAATGTTCTAGAGAGCTGTTTACAGACCTTTTCATCAGCAGGCAAGCAATCTCAGAC 3441
 Db 2102 GCTGAGTACTAGGTGCGAGTAATATAGCTTTTATCAGCAACGGGAAAGAAACCTTATGTT 2161
 QY 3442 AAATCATCAGCAATGGAAGACTGTAAAGCAATGCAAGCAGCAGCTCAATATCCAGCTTCT 3501
 Db 2162 AGAAGCACTGACACTGAGACTGGAAGCAATTTATGTAATTTAGAGCTGAACCTTTCA 2221
 QY 3502 GGAAGATTAAGGGGAGCTTGTATGTTCTGAAATGGATCAGACATGAAATGAAGA 3557
 Db 2222 ACCGAAATGATCAAGAGGAAAGAGTTTCAGAGGACAGAGAACTGGAGTTAGTA 2277
 RESULT 9
 AAS44727
 ID AAS44727 standard; DNA; 714 BP.
 AC AAS44727;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human full-length polynucleotide sequence #152.
 XX
 KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat; mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia; cancer; lymphoma; neuroblastoma; autolimmune disorder; cell proliferation; nervous system disorder; inflammatory disorder; cell differentiation; ds; angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn; genetic disorder; bone regeneration; tendon; ligament; tissue repair; cytoskeletal; antirheumatic; antiarthritic; vulnery; antiinflammatory; antibacterial; immunosuppressive; vasotropic; antiparkinsonian; neuroprotective; osteopathic; antidiabetic; antiallergic; immunostimulant; analgesic; gene therapy.
 KW
 OS Homo sapiens.
 XX
 XX
 PN WO200164834-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04926.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 XX
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PR 17-JUN-2000; 2000US-0597707.
 XX
 PR 14-JUL-2000; 2000US-0616807.
 XX
 PR 19-SEP-2000; 2000US-0664641.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 XX
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
 XX
 PI Dmanac R;
 XX
 DR WPI: 2001-589862/66.
 XX
 DR P-PSDB: AAU27827.
 XX
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries
 XX
 PT prepared from various human tissues, for diagnosis, treatment of
 XX
 PT cancer, neurological, inflammatory disorders and for use in arrays for
 XX
 PT detection -
 XX
 CC Claim 1; SEQ ID NO 152; 153bp; English.
 CC
 CC Sequences AAS44727-AAS44919 represent full-length polynucleotides and
 CC
 CC contig polynucleotides encoding polypeptides of the invention. The DNA
 CC
 CC and protein sequences are useful for the treatment, diagnosis and
 CC
 CC prevention of various types of disorder in a mammalian subject such as a

CC human, dog, monkey, mouse, hamster or rat. The disorders include cancers
 CC such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
 CC as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,
 CC cell proliferation, cell differentiation, stem cell growth factor,
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells
 CC in culture to give rise to neuroepithelial cells that can be used to
 CC augment or replace cells damaged by illness, accidental damage or genetic
 CC disorders. The sequences may also be used for regeneration of bone,
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.
 CC Note: Some sequences for this patent did not form part of the printed
 CC specification, but were obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

CC Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;

Query Match 12.0%; Score 499; DB 22; Length 714;
 Best Local Similarity 100.0%; Pred. No. 6.1e-113;
 Matches 499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3674 AGGAGAGAAATAGTATGATGACTGACTAGTGTGTACAGAGAAAGCAACCTTAAG 3733
 DB 1 AGGAGAGAAATAGTATGATGATGACTGACTAGTGTGTACAGAGAAAGCAACCTTAAG 60
 OY 3734 GCAGTCAGGAAAGTCGAAAGAGAGCCATACGGCTTCAGAAATCTGTGAACAGCAGTGGC 3793
 DB 61 GCAGTCAGGAAAGTCGAAAGAGAGCCATACGGCTTCAGAAATCTGTGAACAGCAGTGGC 120
 OY 3794 CTGAGAGAAAGAGGCTCAAGAAAGATATTTAGAAATAGATGACAGATAGTCCGC 3853
 DB 121 CTGAGAGAAAGAGGCTCAAGAAAGATATTTAGAAATAGATGACAGATAGTCCGC 180
 OY 3854 CAAAAAAGGTTAAAGAGGCCAGCCACCAAAACCTCTGTGTGAGTACACCAAAAGAG 3913
 DB 181 CAAAAAAGGTTAAAGAGGCCAGCCACCAAAACCTCTGTGTGAGTACACCAAAAGAG 240
 OY 3914 AGCCACAATGAAATCTCTAAAGAAAGAAAGAAATCTGTGACCTCCAGCACAG 3973
 DB 241 AGCCACAATGAAATCTCTAAAGAAAGAAAGAAATCTGTGACCTCCAGCACAG 300
 OY 3974 AGGAGAGAAAGAAAGAAAGAAAGAAAGTGAATTCGGAACAGAAAGTCCAAAGCAAC 4033
 DB 301 AGGAGAGAGAAAGAAAGAAAGAAAGTGAATTCGGAACAGAAAGTCCAAAGCAAC 360
 OY 4034 AGCAGCGAGTGTCAAGAGAGAGAGAGAGAGAGATCTCTGTAATCTAGTCAATTG 4093
 DB 361 AGCAGCGAGTGTCAAGAGAGAGAGAGAGAGAGATCTCTGTAATCTAGTCAATTG 420
 OY 4094 AATTCACACAGTTCACACACAGAAAGAGAGAAAGACCATCAAAAGCCATCCAT 4153
 DB 421 AATTCACACAGTTCACACACAGAAAGAGAGAAAGACCATCAAAAGCCATCCAT 480
 OY 4154 CACACCAAAAAAATGT 4172
 DB 481 CACACCAAAAAAATGT 499

RESULT 10
 ABR35358
 ID ABR35358 standard; cDNA: 2496 BP.
 AC ABR35358;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA encoding secreted protein #496.
 XX
 KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;

KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.

OS Homo sapiens.

PN WO200177288-A2.

XX 18-OCT-2001.

XX 29-MAR-2001; 2001WO-US10224.

XX 06-APR-2000; 2000US-195562P.

XX (GENM) GENETICS INST INC.

XX PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 XX PI Gulukota K, Graham JR;

XX DR WPI: 2002-179321/23.

XX PT Five hundred and ninety two polynucleotides derived from a variety of
 XX PT human tissue sources which encode secreted proteins, useful for
 XX PT treating immune deficiencies and disorders such as autoimmune disorders
 XX PT -
 XX Claim 1; Page 328-329; 372pp; English.

CC The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABR34863-ABR35454 represent polynucleotides of the invention.

SO Sequence 2496 BP; 822 A; 430 C; 491 G; 753 T; 0 other;

Query Match 11.8%; Score 493; DB 24; Length 2496;
 Best Local Similarity 65.6%; Pred. No. 3.3e-111;
 Matches 752; Conservative 0; Mismatches 390; Indels 5; Gaps 2;

OY 2387 ATCTTCTCATGATGATGCGCTTCAGGAAAGCAACCTAAACTTGTGGTCCAGAG 2446
 DB 83 ATCTTCTCATGATGATGCGCTTCAGGAAAGCAACCTAAACTTGTGGTCCAGAG 142
 OY 2447 AAGAAATATCTCTGAGACATGCTCAAAATTCAGGCTTAATAATGATGTTGATGCG 2506
 DB 143 AAGAGTTTCCCTGAAAGTACAGCAAGAGTACAGCAATTAACCTTGTGAAGTGGC 202
 OY 2507 TACTTGGATGAAAAATTAATTCACAGTAATTCAGGAATCTTAACTTAATGCTAACCA 2566
 DB 203 TGTGGGTATGAAAAACAACAGTCTAAATCTGCCAATTCACCCCTTGTATTAATCAG 262
 OY 2567 CAATATGATGATGATGAGAGCTTCAGCAGAGAGAGGAAATTAATTAACAGATATGT 2626
 DB 263 CGATGTTGTTATGATGAGGATGACCTGACAGAGCAAAAGAGATCAATTAATGATATGT 322
 OY 2627 CACGTCGAGACTTGCTGCTGGAGTGTATTTGTGAAGCTGGCACAAGAACCTGTACC 2686

Db 323 CTCGGTGGATTAGTGTGCTAGTGGCATATAGAGCTTGTCTCAGGAACTTGTTCAC 382
QY 2687 ATGAATCATCATTTAGAACAAATATACAGTATGTGCATTAGCTATCAACGATGATGCT 2746
Db 383 ATGAATATATACCCAGAACAGTTTACAGCTCTGTGCACCTGTATATATGATGATGCT 442
QY 2747 ATCAAGTAGACAGATGTTGGCCAGAAACCTTCACAAAGCCCTTCCGCTTACGGCTTC 2806
Db 443 ACCAAGTAGACAGATATTTGCTCAGAAAGCTGCATTAAGGCACTTGTGAAGTACTGCTCC 502
QY 2807 CACTTGAATATATGGCAATCTGTGCCCTTGTGCAAAAGATCCTGTAAGAGAGAGAAG 2866
Db 503 CATGGAGATATATGGGATCTTGGCTTGTGTGCCAANATCTGTGAAGAGAGAAGAG 562
QY 2867 CTCATGCTTAGGCAATGTTGGTGAATAATATTAATGTAAAGCGGAGATATGGAAGCACC 2926
Db 563 CACAGCGACAGCAAAAGTTTACTGAAAAAATATCATATCGCAGGGGAATCATATTAAGCAGA 622
QY 2927 ATGACAGCTGTAGTAAAAAATATTTGCTCTTCTCCAGAGATATGTTGCATATACAA 2986
Db 623 ATCCATGGCTACTAGAAATATATATCATCTTGGCTGAAATATGATTCATCATGATGA 682
QY 2987 TTCACCTTTGGCAGCATGACCCAGATTAATGTCAAGTACAGGATATGGAACACTTAAAG 3046
Db 683 TTCACCTGCTAGCCCATGATCCAGATTTACAAAGATTCACAAATGATGTGATCGCTTCGT 742
QY 3047 ATGTAAAGATATGTCTTTGGTGTGTTCTGGAATTTAATGGCTTAAAAATGAATAACA 3106
Db 743 ATATCAAAAGAGCTTATGCTTATGCTGATGTAAGTTTATATGCAAAAGATGAACAATA 802
QY 3107 GTCACGCTTTTATCAGAAAGATGTAGAAATATTTAAACAAACAAAGATGCCCAAGAC 3166
Db 803 GCCATGCTCTTATGAGAAGATGCGACAGAACATCAAGTTAAACAGAGATGCCAGTCTC 862
QY 3167 CAGATGATGCAAAAAATGAATGAATACTGTACACTGTGTGTGATGTGCCATGAAATATCA 3226
Db 863 CAGATGATGCCAAGCAAAATGAATAAGTATATAGATATGATGTGTGCTCTGTGTTA 922
QY 3227 TCATGTCAAAGAGTCTACATACAGTTTGAATTCCTTAAGACCCGCTACTACCAAGCTC 3286
Db 923 TAAATAGTAAAGTCTTGTGCA--ATGCAGATTCCTCCAAAGAGCCAGCTCAATGAA 980
QY 3287 GTTCTCTACCTCAACCTGACAAAGATTCAGTAACACAAATAATATCTGCTCTCTGAAA 3346
Db 981 ATTTTACAC--ACCTGAAAAGGACTTCTGTGAAGATTAAGGTTATTTTCAGAGAGA 1037
QY 3347 TGAATTCATTTTTCACCTCTGAAAAACCTTAAACCAACCAATGTTTAGAGCTGTTACA 3406
Db 1038 CAAGAGTACTTCTGTTAACAGSAAAGCCAAAGCCGTGAGTACTAGTGCAGTAATA 1097
QY 3407 AGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAATTCATCAGAAATGAACTGTAA 3466
Db 1098 AGCCTTATCAGCAAGGGAAGAAACCTTATGTTAGAACACTGCTGAGACTGAGTGA 1157
QY 3467 GCAATGCAAGAGCACTCAATCAACCTCTCTCGAAGAAATTAAGGAGAGCTGTGATA 3526
Db 1158 GCAATATTAATGTAATTCAGAGCTGAACCTTCAACCGAAATGATCAAGGGAACAGA 1217
QY 3527 GTTCTGA 3533
Db 1218 GTTCAGA 1224

RESULT 11
AAK85948
ID AAK85948 standard; DNA; 7943 BP.
XX
AC AAK85948;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234997.
PR 25-SEP-2000: 2000US-0234998.
PR 26-SEP-2000: 2000US-0235484.
PR 27-SEP-2000: 2000US-0235834.
PR 27-SEP-2000: 2000US-0235836.
PR 29-SEP-2000: 2000US-0236327.
PR 29-SEP-2000: 2000US-0236367.
PR 29-SEP-2000: 2000US-0236368.
PR 29-SEP-2000: 2000US-0236369.
PR 29-SEP-2000: 2000US-0236370.
PR 02-OCT-2000: 2000US-0236802.
PR 02-OCT-2000: 2000US-0237037.
PR 02-OCT-2000: 2000US-0237038.
PR 02-OCT-2000: 2000US-0237039.
PR 02-OCT-2000: 2000US-0237040.
PR 13-OCT-2000: 2000US-0239335.
PR 13-OCT-2000: 2000US-0239337.
PR 20-OCT-2000: 2000US-0240960.
PR 20-OCT-2000: 2000US-0241221.
PR 20-OCT-2000: 2000US-0241785.
PR 20-OCT-2000: 2000US-0241786.
PR 20-OCT-2000: 2000US-0241787.
PR 20-OCT-2000: 2000US-0241808.
PR 20-OCT-2000: 2000US-0241809.
PR 01-NOV-2000: 2000US-0244617.
PR 08-NOV-2000: 2000US-0244674.
PR 08-NOV-2000: 2000US-0246475.
PR 08-NOV-2000: 2000US-0246476.
PR 08-NOV-2000: 2000US-0246477.
PR 08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246609.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250160.
PR 01-DEC-2000: 2000US-0250391.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 05-DEC-2000: 2000US-0256719.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251868.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.

PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure: SEQ ID NO 40760; 3071pp + Sequence Listing: English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 7943 BP; 2567 A; 1175 C; 1468 G; 2733 T; 0 other;

Query Match 10.5%; Score 438; DB 22; Length 7943;
Best Local Similarity 85.3%; Pred. No. 2.1e-97;
Matches 541; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 3633 GGAGAAAGCTTAGAGGAGCAAAAAACGCCCTCACAGAACAGGAGGAGAAATTAGCTAT 3692
DB 1 GGAGAAAGCTTAGAGGAGCAAAAAACGCCCTCACAGAACAGGAGGAGAAATTAGCTAT 60
QY 3693 GGATGACTTGACTAGTTGGTACAGGAACAGAACTTAAGCAGTCAGCAAGTCGGA 3752
DB 61 GGATGACTTGACTAGTTGGTACAGGAACAGAACTTAAGCAGTCAGCAAGTCGGA 120
QY 3753 AAGAGCCATACGGCTTCAGAAATCTGATGACAGCAGTGGCTTGAGGAAAGAGCTCA 3812
DB 121 AAGAGCCATACGGCTTCAGAAATCTGATGACAGCAGTGGCTTGAGGAAAGAGCTCA 180
QY 3813 AGAAGTATATTAGAAATGAAGATGACAGAAATAGTCCGCAAAAAAGGTTAAAGAG 3872
DB 181 AGAAGTATATTAGAAATGAAGATGACAGAAATAGTCCGCAAAAAAGGTTAAAGAG 240
QY 3873 CCGACCACCAAAACCTTGTTGGAGGTACACCAAAAGAGGCCAACAATATAAATCTC 3932
DB 241 CCGACCACCAAAACCTTGTTGGAGGTACACCAAAAGAGGCCAACAATATAAATCTC 300
QY 3933 TAAAAAAGAGCAAAAAAAATCTGAGCTTCAGACACAGAGGAGGAGGAAGAAGA 3992
DB 301 TAAAAAAGAGCAAAAAAAATCTGAGCTTCAGACACAGAGGAGGAGGAAGAAGA 360
QY 3993 AAGACAAAGTGAATATGCGAAGCAGAACTCCAAAAGCAACGCCAGGTCTCAAGAG 4052
DB 361 AAGACAAAGTGAATATGCGAAGCAGAACTCCAAAAGCAACGCCAGGTCTCAAGAG 420
QY 4053 AGCAGAGAG----- 4062
DB 421 AGCAGAGAGAGCTAAGCATGTGTAACTCTAACTGATCTGTTCGTTACTATATTATA 480
QY 4063 -----AGAGCAAGATCTCTGA 4079
|||||

Db	481	AATCATATTGTAIGCTATCCACATTTGGGTCTTCCCAGAAAGCAGAGCAGAATCTCTCGA	540
QY	4080	ATCTAGTGCAATTTGAATTCACACAGTCGCACACACAGAAAGAGAGAGACCATCAA	4139
Db	541	ATCTAGTGCAATTTGAATTCACACAGTCGCACACACAGAAAGAGAGAGACCATCAA	600
QY	4140	AAGGCCATCACCATTCACACACCACAAAAAATGTG	4173
Db	601	AAGGCCATCACCATTCACACACCACAAAAAATGTG	634
 RESULT 12 AAV87629 standard; CDNA: 439 BP.			
ID	AAV87629		
XX	AAV87629;		
XX	12-FEB-1999 (first entry)		
DT			
DE	EST clone DY17.		
XX			
KW	Expressed sequence tag; secreted protein; haematopoiesis regulator;		
KM	tissue growth; actinin; inhibin; tumour invasion suppressor; Est; human;		
KW	chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;		
KM	receptor; ligand; anti-inflammatory; tumour inhibitor; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO9845437-A2.		
XX			
PD	15-OCT-1998.		
XX			
PF	10-APR-1998; 98WO-US06956.		
XX			
PR	10-APR-1997; 97US-0837312.		
XX			
PA	(GENE) GENETICS INST INC.		
XX			
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;		
PI	Racie LA, Spaulding V, Treacy M;		
XX			
DR	WPI: 1999-070078/06.		
XX			
PT	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
XX	ovary, pituitary, retina and colon cDNA libraries		
PS	Claim 1; Page 126; 641pp; English.		
XX			
CC	The present sequence represents an expressed sequence tag (EST), and is		
CC	a polynucleotide of the invention. The polynucleotides of the invention		
CC	are all secreted EST sequences isolated from a variety of human tissue		
CC	sources. The EST sequences and proteins encoded by them are predicted to		
CC	have useful biological activities which would make them suitable for		
CC	treating, preventing or ameliorating medical conditions in humans and		
CC	animals, although no supporting data is given. Suggested activities		
CC	include nutritional activity, immune stimulating or suppressing activities,		
CC	haematopoiesis regulating activity, tissue growth activity,		
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory		
CC	activity, cadherin/tumour invasion suppressor activity, tumour inhibition		
CC	activity. The EST sequences are also stated to be useful for gene		
CC	therapy.		
XX			
SQ	Sequence 439 BP; 150 A; 101 C; 85 G; 103 T; 0 other;		
QY			
Query Match	9.9%; Score 411.4; DB 20; Length 439;		
Best Local Similarity	98.6%; Pred. No. 2.1e-91;		
Matches 415; Conservative	0; Mismatches 6; Indels 0; Gaps 0;		
3216 CATGAAATCATCATGCTCAAGAAGACGACACTACATCAAGTTGGAAATCTCTTAAGACCCGGT	3275		
13 CATGGCCCTCATCATGCTCAAGAAGACGACACTACATCAAGTTGGAAATCTCTTAAGACCCGGT	72		

OY		3276	ACTACCAAGCTCGTTTCTTCACCTGACAAGATTTTAGTAACACCATAATTATCT	3333
Dd		73	 ACTACCAAGCTCGTTTCTTCACCTGACAAGATTTTAGTAACACCATAATTATCT	132
OY		3336	GCCGCCGAATGTAAATCATTTTTCACCTCGGGAAAACCTPAAAACAACCAATGTTTAGG	3395
Dd		133	 GCCCTCCTGAAATGTAAATCATTTTTCACCTCGGGAAAACCTPAAAACAACCAATGTTTAGG	192
OY		3396	AGCTGTTAACAAGCACCTTTTCATCAGCAGGCAGAAGCAATTCCTAGACCAAATCATCAGCAAT	3455
Dd		193	 AGCTGTTAACAAGCACCTTTTCATCAGCAGGCAGAAGCAATTCCTAGACCAAATCATCAGCAAT	252
OY		3456	GGAACACGTGAAGCAATGCAAGCAGCAGCTCAAAATCCAAAGCTCTCTGGAGAAATAAAGG	3515
Dd		253	 GGAACACGTGAAGCAATGCAAGCAGCAGCTCAAAATCCAAAGCTCTCTGGAGAAATAAAGG	312
OY		3516	GAGCGTTGATGTGTTCTGTAATGATGCACAGTGAAATGAAGATATACCAATGTTCTTGACC	3575
Dd		313	 GAGCGTTGATGTGTTCTGTAATGATGCACAGTGAAATGAAGATATACCAATGTTCTTGACC	372
OY		3576	TTTTCCGGGGAAAAAAGTGACAAGAGAGACGACTGTGATCTTGTAAAGTCTGAATTGA	3635
Dd		373	 TTTTCCGGGGAAAAAAGTGACAAGAGAGACGACTGTGATCTTGTAAAGTCTGAATCGGA	432
OY		3636	G 3636	
Dd		433	G 433	
RESULT 13				
ID	ABL89669	standard; cDNA; 461 BP.		
XX	ABL89669;			
DT	24-MAY-2002	(first entry)		
DE	Human polynucleotide SEQ ID NO 231.			
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;			
KW	antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antifurc;			
KM	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;			
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;			
KM	neurological disease; infection; human; secreted protein; gene; ss.			
XX	Homo sapiens.			
OS				
XX	WO200190304-A2.			
PN				
XX	29-NOV-2001.			
PD				
XX	18-MAY-2001; 2001WO-US16450.			
PF				
XX	19-MAY-2000; 2000US-205515P.			
PR				
XX	(HUMA-) HUMAN GENOME SCI INC.			
PA				
XX	Birse CE, Rosen CA;			
PI				
XX	WP1; 2002-122018/16.			
DR				
XX	P-PSDB; ABB89260.			
DR				
XX	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and			
PT	prevention of neural, immune system, muscular, reproductive,			
PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative			
PT	disorders -			
XX				
PS	Claim 4; SEQ ID NO 231; 2081bp + Sequence Listing; English.			
XX				
CC	The invention relates to novel genes (ABL89449-ABU90853) and proteins			
CC	(ABB89040-ABB90444) useful for preventing, treating or ameliorating			
CC	medical conditions e.g. by protein or gene therapy. The genes are			

CC isolate from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful.
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed Specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct-sequences.

SQ Sequence 461 BP; 135 A; 84 C; 116 G; 121 T; 5 other;

Query Match 8.5%; Score 353.4; DB 24; Length 461;
Best Local Similarity 98.9%; Pred. NO. 4.6e-77;
Matches 354; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

SQ Sequence 461 BP; 135 A; 84 C; 116 G; 121 T; 5 other,

Query Match	8.5%	Score 353.4	DB 24	Length 461
Best Local Similarity	98.9%	Pred. No. 4.6e-77		
Matches 354	Conservative	0	Mismatches 4	Indels 0
				Gaps 0

QY	1	ATGGCTATTC	AAAGAC	TAGGAC	CAAC	AATATG	AAAAAT	TACATAT	CCCC	TGGGGT	CAAG	60
Db	104	ATGGCTATTC	AAAGAC	TAGGAC	CAAC	AATATG	AAAAAT	TACATAT	CCCC	TGGGGT	CAAG	163
QY	61	GAATATAC	AGATTA	ATCTTA	AAAGAG	GAGATG	GTGAG	ACCATTA	AAAGATG	GGTGTG	GAAA	120
Db	164	GAATATAC	AGATTA	ATCTTA	AAAGAG	GAGATG	GTGAG	ACCATTA	AAAGATG	GGTGTG	GAAA	223
QY	121	ACTTTATG	CATATG	ACAC	GAGACT	CTCTA	AAAGAA	AGAGCTT	ATTATTA	AACTTACT	180	
Db	224	ACTTTATG	CATATG	ACAC	GAGACT	CTCTA	AAAGAA	AGAGCTT	ATTATTA	AACTTACT	283	
QY	181	TTACATCT	GCTTAG	ATTTTT	TTTCTC	ACAGCAT	CTCGT	AAAGATG	TTCGCTT	ACTGTGTA	240	
Db	284	TTACATCT	GCTTAG	ATTTTT	TTTCTC	ACAGCAT	CTCGT	AAAGATG	TTCGCTT	ACTGTGTA	343	
QY	241	GCCGCTG	CCCTG	CTGATAT	TTTACG	AGTTAT	GCTCCT	CGAAGCT	CCCTTAC	CATCCCT	300	
Db	344	GCCGCTG	CCCTG	CTGATAT	TTTACG	AGTTAT	GCTCCT	CGAAGCT	CCCTTAC	CATCCCT	403	
QY	301	GATTAAC	TAAAGAT	ATATTT	ATGTTT	ATAC	CAAGAC	AGTTGA	AGGGCT	AGAGGATA	358	
Db	404	GATTAAC	TAAAGAT	ATATTT	ATGTTT	ATAC	CAAGAC	AGTTGA	AGGGCT	AGAGGATA	461	

RESULT 14

ID ABS56684 standard; cDNA; 1104 BP.

AC ABS56684 ;

DT	23-JAN-2003	(first entry)
YY		

DE
XX

Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.

kw	Euchromosome	fragile	intelligence	delay	syndrome	protein 20.02; human;
kw	Euchromosome	fragile	intelligence	delay	syndrome	protein 20.02; human;
kw	Euchromosome	fragile	intelligence	delay	syndrome	protein 20.02; human;

aa
OS Homo sapiens.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	
		184..732

FT	/product=	"euchromosome fragile intelligence delay
FT		syndrome-associated protein, 20.02"

XX 29-MAY-2002.

PF 26-OCT-2000; 2000CN-0125797.

XX 26-OCT-2000; 2000CN-0125797.
PR
XX
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI

CC This invention describes the novel human euchromosome fragile
CC intelligence delay syndrome associated protein, 20.02. The polypeptide is
CC used in treating diseases such as euchromosome fragile intelligence delay
CC syndrome, and dementia. This sequence encodes the human euchromosome
CC fragile intelligence delay syndrome-associated protein, 20.02 described
CC in the disclosure of the invention

SQ Sequence 1104 BP; 362 A; 188 C; 228 G; 326 T; 0 other;

Query Match	8.48;	Score 350.4;	DB 24;	Length 1104;
Best Local Similarity	69.7%;	Pred. No. 3.7e-76;		
Matches 488; Conservative	0;	Mismatches 211;	Indels 1;	Gaps 1

QY	104	ACGAAAGACCTTAACAGAGTATCTTAAAGTAGAGTCACATGACCCTGAGAACCTCTTTG	11
Db	1	ACGGAAGATCTCACAGAAATATTTAAAGGTATGATCATGATTCACAGAAAGCTTTGG	60
QY	1101	ACATGATGTTATTTGTGTCATATAGTTCACGCTGCTAAAAAGGATATTTCTGTGCATGA	1166
Db	61	TCATGATGTCATTTGTACTATATAATTAACGCTGCCAAGAGGAGACCTGGCT-TACTAATGA	119
QY	1161	TCACCTACTTAATTTTGTGAGAGAGAACTTTGACAAAGATGGAGATATAGCAAAAG	1221
Db	120	TCAGCTGCTGGCTTTGTATAGGGAAGAACACTGATTAACCGTGGCGAATAGAAAGAA	179
QY	1221	AGCCATGATGGACCTTGGCCAAATTTATTAAGAAATATGCTTACAGCTCAGCAGCTGGAA	1281
Db	180	ACGTATGATGGGTCTGGCCTCAGCTTTATTAAGAAATATCTGCTTCATGTGTGAACAGGAA	239
QY	1281	AGATGTGCAAAACAGATATGATATGATCAAGACAAATTCGTCATATATATATGCAAA	1341
Db	240	GGAGCTGCGAGGAAAGTCACGCTGGATTAAGAGCAAACTTCGTCATATTTATATACAA	299
QY	1341	TGATATGTGATGTCACACTACTTGTGTGAACGGATCTTGCTCAATACATGGTCTCCACA	1401
Db	300	CAGCANTAGCGCAAAATTTGTTGGTATGAGAAATCTTTTGTCAGATCTCTGCCCCACA	359
QY	1401	TTTTAGAACCTACAGAACGGATGAATGCTTATATTACTTGTATGCCACACTGATTTAAA	1461
Db	360	CCTGGAAACAGAAAGAGAAATGAATGCTTATATTACTTTATGCTAGTTGGATCCAAA	419
QY	1461	TGCTGTGAAGCATTTGATGAATATGTGAAATGTCAAAATCTGCTCCGACATCAAGTAAA	1521
Db	420	TGCTGTAAAGCTCTCAACGAAATGTGGAATGTGCACAAATGCTTTGGAGCCATATACG	479
QY	1521	GGATTTGCTTGACTTGATTAAGCAACCCAAACAGATGCCAGTGTCAAGGCCATATTTTC	1581
Db	480	CGAAGCATTTGATTTGGACAAGCAGCCCTCATCATCAGAGGCTAATCGTTCTCCATGTTGG	539
QY	1581	AAAAGTGATGGTTATTAACAAGAAATTTACCTGATCTCTGTATAGGCTCAGATTTTCATGA	1641
Db	540	AAAACGTAGACCAATAGCAAAAGAAATTTGCTGTAGCCCGGGAAGAGCAAAATTTTGTGA	599
QY	1641	GAATTTACACACAGGTGTAGAAGATGATGAGAAATATAGAAGAACAGCTTTAGAACTTGT	1701
Db	600	GAATTTAACACAGGTCTCGCGGATGATGAGAAACCTTCGGCTCAGTTGAGATTAATAT	659
QY	1701	TATGTCACATGCTCTCTGCAAAGCAGGCTGAAGGTTGTGTG 1740	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 08:52:32; Search time 14865 Seconds

(without alignments)
11484.418 Million cell updates/sec

Title: US-09-512-581B-3

Perfect score: 4173

Sequence: 1 atgctcattcaagaactag.....cacaacacaaaaaatgtg 4173

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_inv:*
33: em_hlg_other:*
34: em_hlg_mus:*
35: em_hlg_pln:*
36: em_hlg_rod:*
37: em_hlg_mam:*
38: em_hlg_vrt:*
39: em_hlg_hum:*
40: em_hlg_mus:*
41: em_hlg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4165.6	99.8	5309	9	AB023196	AB023196 Homo sapi
2	4165.6	99.8	7444	9	HS26H231	AL137201 Novel hum
3	4141	99.2	5253	9	HS095825	U95825 Human andro
4	3471.8	83.2	5317	10	AY102267	AY102267 Mus muscu
5	2816.6	67.5	5944	10	AK122414	AK122414 Mus muscu
6	1815.6	43.5	6744	9	AF294791	AF294791 Homo sapi
7	1754.4	42.0	1852	9	HS050533	U50533 Human BRCA2
8	1464.6	35.1	4252	9	BC039256	BC039256 Homo sapi
9	1041.4	25.0	5177	9	AB014548	AB014548 Homo sapi
10	937.6	22.5	1976	10	BC021408	BC021408 Mus muscu
11	932.2	22.3	2265	9	AK098331	AK098331 Homo sapi
12	932.2	22.3	2705	9	BC041361	BC041361 Homo sapi
13	447	10.7	2079	9	AK026889	AK026889 Homo sapi
14	447	10.7	137246	9	HS49010	284572 Human DNA S
15	447	10.7	168487	9	AL138820	AL138820 Human DNA
16	443	10.6	72157	2	AC016449	AC016449 Homo sapi
17	411.4	9.9	439	6	BD059747	BD059747 Secretd
18	389.8	9.3	54398	2	AC068224	AC068224 Homo sapi
19	332.4	8.0	236508	2	AC111126	AC111126 Mus muscu
20	292	7.0	295	6	BD026562	BD026562 Sequence
21	288.6	6.9	263479	2	AC125986	AC125986 Rattus no
22	288	6.9	3779	9	BC009650	BC009650 Homo sapi
23	232.2	5.6	772	6	BD159124	BD159124 Primer fo
24	232.2	5.6	2212	9	AK021757	AK021757 Homo sapi
25	232.2	5.6	227	6	AX494307	AX494307 Sequence
26	208.8	5.0	227	6	AC068224	AC068224 Homo sapi
27	207.4	5.0	113704	9	HS267P19	275889 Human DNA s
28	206.8	5.0	120652	9	AL353724	AL353724 Human DNA
29	206.8	4.3	85592	2	AL358892	AL358892 Mus muscu
30	179.6	4.3	179132	10	AL358892	AL358892 Mouse DNA
31	179.6	4.3	236508	2	AC111126	AC111126 Mus muscu
32	179.6	4.3	236508	2	AC094469	AC094469 Rattus no
33	173.2	4.2	110000	2	AC110493	AC110493 Mus muscu
34	142.8	3.4	184321	2	AC112263	AC112263 Mus muscu
35	142.8	3.4	190058	2	AC116574	AC116574 Mus muscu
36	142.8	3.4	298053	2	AC095342	AC095342 Rattus no
37	142.6	3.4	233589	2	AC113150	AC113150 Homo sapi
38	131	3.1	98856	2	AC068352	AC068352 Homo sapi
39	131	3.1	145450	2	AC022862	AC022862 Homo sapi
40	131	3.1	160962	2	AC090081	AC090081 Homo sapi
41	131	3.1	167671	2	AC087748	AC087748 Homo sapi
42	124	3.0	69740	2	AC068352	AC068352 Homo sapi
43	121.4	2.9	145450	2	AC023006	AC023006 Homo sapi
44	121.4	2.9	162775	2	AC007475	AC007475 Drosophila
45	119.2	2.9	173613	3		

ALIGNMENTS

RESULT 1
AB023196 5309 bp mRNA linear PRI 10-MAY-2002
LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.
DEFINITION AB023196
ACCESSION AB023196
VERSION AB023196.2 GI:20521717
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M.,
Miyaajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.

Pred. No. is the number of results predicted by chance to have a

XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
 JOURNAL DNA Res. 6 (1), 63-70 (1999)
 MEDLINE 99246063
 PUBMED 10231032
 REFERENCE 2 (bases 1 to 5309)
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
 JOURNAL Direct Submission
 Submitted (04-FEB-1999) Otsu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)
 On May 9, 2002 this sequence version replaced gi:4589601.

COMMENT
 FEATURES
 source

1..5309
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="hj07056s1"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pbluescriptII SK plus"
 /dev_stage="adult"
 /note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 722 was derived from Fj09522 and 723 - 5309 was derived from hj07056)."
 1..5309
 /gene="KIAA0979"
 <1..4452
 /gene="KIAA0979"
 /codon_start=1
 /product="KIAA0979 protein"
 /protein_id="BA076823.2"
 /db_xref="gi:20521718"

gene
 CDS

/translation="GGGRRRLDAAAADPGPSSRSSEATRRGRNLSVMAHSKTR
 NDGKITPPGKKEISDKISKEMVRRLKMYKTPMDMDDESEKELYNLALASD
 FEKHPKDVRLVACCLADIFRLYAEAPITSPDKLIDIMFTTROLGLDEKSPQ
 FNRFLYLENLAWKSYNCFELEDNEAFITQRLTFLSVNNHNOVHMHMADLMS
 SIICEDTQSELDLIVLVNPAHKNKQADLAKALAKRTQAILEPTTFNFQV
 LMGRKTSIDSEHPDLILELYNIDSHLIDSVLPOLFEKLKSDNENRLOVYLAK
 MFGAKSEIASONKLMOCYIGRNDIHPIRECVKFAASHCLMNHDLAKDLLEYLK
 VRSHDPEAIRHDIVSYVYAAKDILVNDHLNFRERILDKRMRYKRAMGLAO
 IYKTAQSAAGKDAKAOIAWKDKLHITYQNSIDRLVERLFAQYVPHNLETTE
 RMKCLYLYLAVTLDNAAKALNEMKCONLHQVLDLILKOPTDAVYAFISKVA
 VITRNLPDPCAODPMFKFTQVLEDDERIKRQLEVLVSPSCQAEGCAVEIKKIG
 NPQPTNPLEMIKFLERLAPVHIDTESIALIKOVNKSIDGAEDEDEVPTDOAI
 RAGLELKVLSTPTSPISFSAETPESILLACKMDKEVAAALDIFRNTGSKTEEDP
 HIRSAFLPVLHKSKKGPPIROAKTAIHCITAHIESKETQFAQIEPLHKSIDPNSLH
 LITPLVIGHIALAPQFAPALPSLAVTFIVKDLMDRLPGKTKLWLPDEVSP
 ETWVKIQAIKMNVNWLGMNNHSGSTSLRLTTLIHSGDLTEQGLSKPMSRL
 FLAAGSAIVKLAQPCYHEIITTEYOYLCALAINDECYQVAVOAKLHKSILRLP
 LEVYALCALCAKDPKERRBAHARQCLVKNINVRREYIKOHAAVEKLSILPEVVY
 TIHLAHPDYVYQDIEQLKDYECLEPVEILIMAKNENSHAFIRKMYENIKOTKD
 AAGPDDAKMKEKITVCDVANIIMSKSTYLSLSPKDPVLPAPFETQPDNFSNTN
 YLPPEKMSFTTPGPKPTNVIGAVNKLPSAGKQSQKSSMETVSNASSSSNSPG
 RIKGRLDSEMDHEDYWSPLPKGKSDKRDSDLVRSLEKPGRKPTPYEDQ
 EKLGMDLTKLVQOKPKGSORSRKRGHRTASEDOQMPKRLKEDILNEDHONSP
 PKKRGRRPPKLGSGTPKKEPTMTKTKGSKKSGPAPPEEEROSGNTPOKSK
 SKQHRVRRKQAQRAESPSSAIESTOSTPOKGRKRPKTESPSPKPKVVRGSKQA
 TKENDSSEEDVFOGSSPVDDIPEETEEEBEVSTVNVRRRSKERRR"

BASE COUNT 1774 a 955 c 1115 g 1465 t
 ORIGIN

Query Match 99.8% Score 4165.6; DB 9; Length 5309;
 Best Local Similarity 99.9% Pred. No. 0; Mismatches 4; Indels 0; Gaps 0;
 Matches 4168; Conservative 0;

QY 1 ATGGCTATTCAGAGACTAGAGCAATGATGGAATAATTCATATCCGCTGGGCTCAAG 60
 Db 109 ATGGCTATTCAGAGACTAGAGCAATGATGGAATAATTCATATCCGCTGGGCTCAAG 168
 QY 61 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGAGACAGATTAAAGATGGTTGTGAAA 120

Db 169 GAAATATCAGATTAATAATATCTAAAGAGAGATGGTGAGACAGATTAAAGATGGTTGTGAAA 228
 QY 121 ACTTTATGATATGAGACCAAGAGACTGTGAGAAAGAAAAGAGCTTTTAAACTAGCT 180
 Db 229 ACTTTATGATATGAGACCAAGAGACTGTGAGAAAGAAAAGAGCTTTTAAACTAGCT 288
 QY 181 TTACATCTGCTGATGATATTTTCTCAAGAGCTCTGTAAAGATGTTCCGTACTGTA 240
 Db 289 TTACATCTGCTGATGATATTTTCTCAAGAGCTCTGTAAAGATGTTCCGTACTGTA 348
 QY 241 GCGTGTGCTGCTGATATTTTCAAGATTTATGCTCTCGAAGCTCTTACACATCCCT 300
 Db 349 GCGTGTGCTGCTGATATTTTCAAGATTTATGCTCTCGAAGCTCTTACACATCCCT 408
 QY 301 GATAACTGAAGGATATATTTATGCTTATACAGACAGCTGAAGGGGCTGAGATACA 360
 Db 409 GATAACTGAAGGATATATTTATGCTTATACAGACAGCTGAAGGGGCTGAGATACA 468
 QY 361 AAGAGCCCAATTCATATGATATTTTATTTACTTGAAGACATTTGCTGGTCAAGTCA 420
 Db 469 AAGAGCCCAATTCATATGATATTTTATTTACTTGAAGACATTTGCTGGTCAAGTCA 528
 QY 421 TATACATATGCTTTGATTAAGATATGCAATGAATTTTACCCAGCTATACAGAAC 480
 Db 529 TATAACATATGCTTTGATTAAGATATGCAATGAATTTTACCCAGCTATACAGAAC 588
 QY 481 TTATTTTCACTTATTAACCAATGGCCACAAATGAGAAAGCCATATGCAATGAGACTT 540
 Db 589 TTATTTTCACTTATTAACCAATGGCCACAAATGAGAAAGCCATATGCAATGAGACTT 648
 QY 541 ATGAGCTCTATTTATTTGTAAGAGTATATACAGTGTCTCAGAGCTTTTGGATACGTTT 600
 Db 649 ATGAGCTCTATTTATTTGTAAGAGTATATACAGTGTCTCAGAGCTTTTGGATACGTTT 708
 QY 601 GTAATCTGTAAGCTGCTCATTAAGATTTTAAACAAGAGATATGATTTGGCAAAGCT 660
 Db 709 GTAATCTGTAAGCTGCTCATTAAGATTTTAAACAAGAGATATGATTTGGCAAAGCT 768
 QY 661 TTACTGAAGAGAGAGAGCTCAAGCTATTTAGAGCAATATATACCATTTTAAACAGATT 720
 Db 769 TTACTGAAGAGAGAGAGCTCAAGCTATTTAGAGCAATATATACCATTTTAAACAGATT 828
 QY 721 CTGATGCTTGGGAAACATCTATCAGAGGATTTGTCAAGAGATGCTTTGACTTAATTT 780
 Db 829 CTGATGCTTGGGAAACATCTATCAGAGGATTTGTCAAGAGATGCTTTGACTTAATTT 888
 QY 781 GAGCTCTACATATTTGATATGATATTTGCTGCTCTGCTTTTAAACCCAGCTTGAATTT 840
 Db 889 GAGCTCTACATATTTGATATGATATTTGCTGCTCTGCTTTTAAACCCAGCTTGAATTT 948
 QY 841 TTAAGAGCAATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 900
 Db 949 TTAAGAGCAATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1008
 QY 901 GGGGCAAAAGATTCAGATTTGGCTTCTCAAAACAGACACTTTGGCAGTGTACTTGGGC 960
 Db 1009 GGGGCAAAAGATTCAGATTTGGCTTCTCAAAACAGACACTTTGGCAGTGTACTTGGGC 1068
 QY 961 AGGTTTAATGATATTCATATGATATGATATGATATGATATGATATGATATGATATG 1020
 Db 1069 AGGTTTAATGATATTCATATGATATGATATGATATGATATGATATGATATGATATG 1128
 QY 1021 CTCATGAACCATTCGATTTTGAAGAGACTTAACAGAGTATCTTAAAGTGAAGTCAAT 1080
 Db 1129 CTCATGAACCATTCGATTTTGAAGAGACTTAACAGAGTATCTTAAAGTGAAGTCAAT 1188
 QY 1081 GACCCCTGAGAGAGTATTTAGACATGATGTTTGTGCAATATGATTTAGAGCTGCTAAAAG 1140
 Db 1189 GACCCCTGAGAGAGTATTTAGACATGATGTTTGTGCAATATGATTTAGAGCTGCTAAAAG 1248
 QY 1141 GATATCTTCTGTCATATGATCACTTACTTAATTTTGTGAGAGAGAGAACTTGAACAAA 1200

Db 1249 GATATTCTTGCTCAATGATCACTTACTTAATTTTGTGAGAGAGAAACATTAGACAAA 1308
QY 1201 CGATGGAGATACGCAAGAGCCATGATGGGACTGCCCAATTTATAGAAATATGCT 1260
Db 1309 CGATGGAGATACGCAAGAGCCATGATGGGACTGCCCAATTTATAGAAATATGCT 1368
QY 1261 TTACAGCTCAGAGCTGGAAAAGATGCTGCAAAACAGATAGCATGATCAAAAGCAATTTG 1320
Db 1369 TTACAGCTCAGAGCTGGAAAAGATGCTGCAAAACAGATAGCATGATCAAAAGCAATTTG 1428
QY 1321 CTACATATATATATCAAAAAATAGTATGTATGATGCACTCTTGTTAGAGGATCTTGCT 1380
Db 1429 CTACATATATATATCAAAAAATAGTATGTATGATGCACTCTTGTTAGAGGATCTTGCT 1488
QY 1381 CAATACATGGTTCTTCACAATTTAGAACTACGAAACGGATGAAATGCTTATATTTCTTG 1440
Db 1489 CAATACATGGTTCTTCACAATTTAGAACTACGAAACGGATGAAATGCTTATATTTCTTG 1548
QY 1441 TATGCCACACTGGATTTAAATGCTGGAAGCACTTGATGAAATGTGAAATGTCAAAAT 1500
Db 1549 TATGCCACACTGGATTTAAATGCTGGAAGCACTTGATGAAATGTGAAATGTCAAAAT 1608
QY 1501 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
Db 1609 CTGCTCCGACATCAAGTAAAGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1668
QY 1561 AGTGCAAGGCCATATTTTCAAAAGTGATGGTTATTTACAAAGAAATTTACTGATCTGCT 1620
Db 1669 AGTGCAAGGCCATATTTTCAAAAGTGATGGTTATTTACAAAGAAATTTACTGATCTGCT 1728
QY 1621 AAGGCTCAGAGATTTTCATGAGAAATTCACACAGGTTGTAGAGATGATGAGAAATTAAG 1680
Db 1729 AAGGCTCAGAGATTTTCATGAGAAATTCACACAGGTTGTAGAGATGATGAGAAATTAAG 1788
QY 1681 AAGCACTTGAAGTACTTGTATGTCACACATGCTCCTGCAGAGGCTGAAGTTGTGTC 1740
Db 1789 AAGCACTTGAAGTACTTGTATGTCACACATGCTCCTGCAGAGGCTGAAGTTGTGTC 1848
QY 1741 CGTGAATTAACATAAGATTTGGGCAACCCCAACAGCCTACAAATCCTTCCGGAATG 1800
Db 1849 CGTGAATTAACATAAGATTTGGGCAACCCCAACAGCCTACAAATCCTTCCGGAATG 1908
QY 1801 ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCACATAGATACCAGATCTATCACTGCT 1860
Db 1909 ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCACATAGATACCAGATCTATCACTGCT 1968
QY 1861 CTTATTAAACAAGTGAACAATCAATAGATGGAACAGCAGATGATGAAGTGAAGGCTTT 1920
Db 1969 CTTATTAAACAAGTGAACAATCAATAGATGGAACAGCAGATGATGAAGTGAAGGCTTT 2028
QY 1921 CCAACTGATCAAGCATCGACAGGCTTGAAAGCTTAAAGCTTACCTCATTTTACAT 1980
Db 2029 CCAACTGATCAAGCATCGACAGGCTTGAAAGCTTAAAGCTTACCTCATTTTACAT 2088
QY 1981 CCCATCTCATTTCTATCTGCTGAACCATTTGAATCACTACTGCTGTCTGAAAATGAT 2040
Db 2089 CCCATCTCATTTCTATCTGCTGAACCATTTGAATCACTACTGCTGTCTGAAAATGAT 2148
QY 2041 GATGAAAAAGTGAAGCAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTTGAA 2100
Db 2149 GATGAAAAAGTGAAGCAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTTGAA 2208
QY 2101 GAGGATTTTCCACACATCGATCAGGCTTGCTTCCGTTTACATCACAAATCTTAAAAA 2160
Db 2209 GAGGATTTTCCACACATCGATCAGGCTTGCTTCCGTTTACATCACAAATCTTAAAAA 2268
QY 2161 GGACCCCCCGTCAAGCCAAATATGCACTTCAATTTGATCCATGCGATATTTTCTAGTAAA 2220
Db 2269 GGACCCCCCGTCAAGCCAAATATGCACTTCAATTTGATCCATGCGATATTTTCTAGTAAA 2328
QY 2221 GAGACCCAGTTGCAACAGATATTTGAGCTCTGCAATTAAGAGCTAGATCCAAAGCACTG 2280
Db 2329 GAGACCCAGTTGCAACAGATATTTGAGCTCTGCAATTAAGAGCTAGATCCAAAGCACTG 2388

QY 2281 GAACATCTCATACACCATTTGTTACTATTGGTCATATTGCTCTCCGTCACCTGATCAA 2340
Db 2389 GAACATCTCATACACCATTTGTTACTATTGGTCATATTGCTCTCCGTCACCTGATCAA 2448
QY 2341 TTTGCTGCTCTCTTGGAATCTTGGTATGCTACTTCTTATGTTGAAAGATCTTCCATGAAT 2400
Db 2449 TTTGCTGCTCTCTTGGAATCTTGGTATGCTACTTCTTATGTTGAAAGATCTTCCATGAAT 2508
QY 2401 GATGCGCTTCAGGAAAAAAGACACTTAACTTTGGGTTCCAGATGAAGATATCTCCT 2460
Db 2509 GATGCGCTTCAGGAAAAAAGACACTTAACTTTGGGTTCCAGATGAAGATATCTCCT 2568
QY 2461 GAGACATGGTCAAAATTCAGGCTATTTAAAAATAGGTTGCGATGAGGCTACTTGGAAATGAA 2520
Db 2569 GAGACATGGTCAAAATTCAGGCTATTTAAAAATAGGTTGCGATGAGGCTACTTGGAAATGAA 2628
QY 2521 AATAATCACAGTAAATCAGAACTTCTACCTTAAGATTGCTAACAACAATATTTGCATAGT 2580
Db 2629 AATAATCACAGTAAATCAGAACTTCTACCTTAAGATTGCTAACAACAATATTTGCATAGT 2688
QY 2581 GATGAGACTTGACAGAACAGGGAAAAATTTAGTAAACAGATATGTCAGCTTGACACTT 2640
Db 2689 GATGAGACTTGACAGAACAGGGAAAAATTTAGTAAACAGATATGTCAGCTTGACACTT 2748
QY 2641 GCTGCTGGAGGCTATTTGGAAGCTGGCAGAACCCCTGTTACGATGAAATCATACCA 2700
Db 2749 GCTGCTGGAGGCTATTTGGAAGCTGGCAGAACCCCTGTTACGATGAAATCATACCA 2808
QY 2701 TTGAAACAATATACGCTATGTCATTAAGTATACAGATGAATGCTATCAAGTAAAGCAA 2760
Db 2809 TTGAAACAATATACGCTATGTCATTAAGTATACAGATGAATGCTATCAAGTAAAGCAA 2868
QY 2761 GTGTTTGCCAGAACTTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAGTATAG 2820
Db 2869 GTGTTTGCCAGAACTTTCACAAAGGCTTTCCGTTTACGGCTTCCACTTGAGTATAG 2928
QY 2821 GCAATCTGCGCCCTTGTGCAAAAGATCCGTAAAGAGAGAAAGACCTCATGCTAGGCAA 2880
Db 2929 GCAATCTGCGCCCTTGTGCAAAAGATCCGTAAAGAGAGAAAGACCTCATGCTAGGCAA 2988
QY 2881 TGTTTGGTGAATAATTAATATGAAGCGGAGATCTGAAGCAGATGCAGCTGTAGT 2940
Db 2989 TGTTTGGTGAATAATTAATATGAAGCGGAGATCTGAAGCAGATGCAGCTGTAGT 3048
QY 2941 GAAAAATTTATGCTCTTCTTACAGAGTATGTTGTCATATTCATTTCACTTTTGCGCA 3000
Db 3049 GAAAAATTTATGCTCTTCTTACAGAGTATGTTGTCATATTCATTTCACTTTTGCGCA 3108
QY 3001 CATGACCCAGATTATGTCAAAAGTACAGATATTTGAACAACCTTAAAGATGTTAAAGATGT 3060
Db 3109 CATGACCCAGATTATGTCAAAAGTACAGATATTTGAACAACCTTAAAGATGTTAAAGATGT 3168
QY 3061 CTTTGCTTGTCTGAAAAATTTAATAGCTTAAAAATGAAGTAAACAGTCAAGCTTTTATC 3120
Db 3169 CTTTGCTTGTCTGAAAAATTTAATAGCTTAAAAATGAAGTAAACAGTCAAGCTTTTATC 3228
QY 3121 AGAAGATGCTAGAAAAATTTTAAACAACAAAAAGATGCCAAGGACCAATGATGCAAAA 3180
Db 3229 AGAAGATGCTAGAAAAATTTTAAACAACAAAAAGATGCCAAGGACCAATGATGCAAAA 3288
QY 3181 ATGAATGAAAACTGTACACTGCTGTGATGTTGCCATGAATATCTCATGTCGCAAGAGT 3240
Db 3289 ATGAATGAAAACTGTACACTGCTGTGATGTTGCCATGAATATCTCATGTCGCAAGAGT 3348
QY 3241 ACTACATACAGTTTGGAAATCTCTTAAAGACCGGATACTACAGGCTGTTCTTCACTCAA 3300
Db 3349 ACTACATACAGTTTGGAAATCTCTTAAAGACCGGATACTACAGGCTGTTCTTCACTCAA 3408
QY 3301 CCTGACAAAGATTTTCAGTAAACCAAAATTTATCTGCGCTCTGAAATGAAATCATTTTTC 3360
Db 3409 CCTGACAAAGATTTTCAGTAAACCAAAATTTATCTGCGCTCTGAAATGAAATCATTTTTC 3468

OY		3361	ACTCTGGAAAACCTTAAACAACCAATGTTCTAGAGCGTGTTAACCAACCCTTCATCA	3422
Dd		3469	ACTCTGGAAAACCTTAAACAACCAATGTTCTAGAGCGTGTTAACCAACCCTTCATCA	3528
OY		3421	GCAGGAAGAATCTCAGACCCAAATCATCACGAATGMAAAGTAAGCAATGCAAGCAGC	3480
Dd		3529	GCAGGAAGAATCTCAGACCCAAATCATCACGAATGMAAAGTAAGCAATGCAAGCAGC	3588
OY		3481	AGCTCAATTCCAAGCTCCCTGTGMAAATAAAGGGAGGCTTGATAGTTCGMAATGGAT	3540
Dd		3589	AGCTCAATTCCAAGCTCCCTGTGMAAATAAAGGGAGGCTTGATAGTTCGMAATGGAT	3648
OY		3541	CACAGTGAATAATGAAGATTACACAATGTCTTCACCTTTGCCGGGAAAAAAGTAGCAG	3600
Dd		3649	CACAGTGAATAATGAAGATTACACAATGTCTTCACCTTTGCCGGGAAAAAAGTAGCAG	3708
OY		3601	AGAGACGCCTGTGATCTTTGAAGTCTGAATTTGAGAAACCTTAGAGCGAGAAAAAAAAAGC	3660
Dd		3709	AGAGACGCCTGTGATCTTTGAAGTCTGAATTTGAGAAACCTTAGAGCGAGAAAAAAAAAGC	3768
OY		3661	CCCCTCACAAGAAACGAGAGAGAAATTTAGTATGGATGTGACTGAATTTGTCAGGAA	3720
Dd		3769	CCCCTCACAAGAAACGAGAGAGAAATTTAGTATGGATGTGACTGAATTTGTCAGGAA	3828
OY		3721	CAGAAACCTTAAAGCGAGTCAAGCGAAGTGGAAAAAGAGCCATAGCGCTTCAGAACTTGAT	3780
Dd		3829	CAGAAACCTTAAAGCGAGTCAAGCGAAGTGGAAAAAGAGCCATAGCGCTTCAGAACTTGAT	3888
OY		3781	GAAACGACGTGGCTTGANGAAAAAAGAGTCTCAAAGAAATATTTAGAAATGGAAGATGAA	3840
Dd		3889	GAAACGACGTGGCTTGANGAAAAAAGAGTCTCAAAGAAATATTTAGAAATGGAAGATGAA	3948
OY		3841	CAGATATAGTCCGCCAAAAGAGGTATAAAGAGAGCCGACCAACAAAACCTCTGTGTGAGGT	3900
Dd		3949	CAGATATAGTCCGCCAAAAGAGGTATAAAGAGAGCCGACCAACAAAACCTCTGTGTGAGGT	4008
OY		3901	ACACCAAAAGAAAGACCCACAATGAAAACCTTTAAAAAAGBACAAAIAAAATCTTGA	3960
Dd		4009	ACACCAAAAGAAAGACCCACAATGAAAACCTTTAAAAAAGBACAAAIAAAATCTTGA	4068
OY		3961	CCTCCAGACACAGAGAGAGAGAAAGAAAGAAACAAAGTCGAAATFAGGAACAGAG	4020
Dd		4069	CCTCCAGACACAGAGAGAGAGAAAGAAAGAAACAAAGTCGAAATFAGGAACAGAG	4128
OY		4021	TCCAAAAGCAAAACAGCACCGAGTGTCAAGAGAGACACAGACGAGAGCAAAATCTCTGAA	4080
Dd		4129	TCCAAAAGCAAAACAGCACCGAGTGTCAAGAGAGACACAGACGAGAGCAAAATCTCTGAA	4188
OY		4081	TCTATGTGAATTGATTCACACAGTCCACACACAGAAAGGACAGAGAAAGCAATCAAAA	4140
Dd		4189	TCTATGTGAATTGATTCACACAGTCCACACACAGAAAGGACAGAGAAAGCAATCAAAA	4248
OY		4141	ACGCCATCACCATCACACCAAAAAAAAAAATGT	4172
Dd		4249	ACGCCATCACCATCACACCAAAAAAAAAAATGT	4280
RESULT 2 HS26H231				
LOCUS		HS26H231	7444 bp mRNA linear PRI 25-JAN-2000	
DEFINITION		Novel human gene mapping to chromosome 13.		
ACCESSION		AL137201		
VERSION		ALI37201.1 GI:6759511		
KEYWORDS				
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarcheria; Primates; Catarrhini; Hominiidae; Homo. Rhodes S. and Huckle,E. Direct Submission Submitted (13-Jan-2000) E-mail contact: humquery@sanger.ac.uk This cDNA sequence was assembled from public domain ESTs and single		
AUTHORS				
JOURNAL				
COMMENT				

pass sequencing reads from expressed DNA templates, aligned to the genomic DNA sequence from the bacterial clones 26H23 (284467), 267P19 (275689) and 43010 (284572).

The EST sequences listed match this sequence with an identity of at least 95% between the coordinates shown.

Further information can be found at <http://www.sanger.ac.uk/NGP/chr13/>

Sanger Centre name : 26H23.Cl13.1

source	location/Qualifiers
exon	1..7444 'organism="Homo sapiens" 'mol_type="mRNA" 'db_xref="taxon:9606" 'chromosome="13" 'map="13q12.3" 1..107
exon	/number=1 108..234
exon	/number=2 127..4470
CDS	/codon_start=1 /product="hypothetical protein" /protein_id="CAB6911.1" /db_xref="gi:6759512" /translation="MAHSKTRTNDGKITTPYGVKKEISDKISKEEMVRLKMYKTFEND MODDEEKEELLYNLALHLASDFLEFKHPDKDVRLLVACCLADIRIYAPAEAPYSPDK LKIDIMFTPIRLQKGLDEPKSPQPNRYPELLNIAWVSYNCEPELDESNLEPQYRT LESYVNNHNOKVHMHVMDLSSITCEBDPYSOELDTDYLVNLYIPARKNLKQAYDLA KALRLKQALEPPTITTFNNOYLMIGKTSISDLSEHVFDTLLELYNTDISHLLSVLPQ LEFKLSNDNERLQVVKLAKMEGAKDEELASONSPLMOCYLGREFNDIVPRLIECV KRAECLKNHHPDLADLDEYELKVRSHDPEEALRGDVAYSITYTAKKDIILVNHLLNF VSEIRLLDNKRVKRAMMGLAOIYKKYALQSAAGKDAKQIAJMKDKLHHYYONSID DLLEVERIAQVMPHNLLETTERMKCEIYLVAFTDILNVKRLNEMKCONLIRQVYD LIDLKQKOTASVKAITSKYMTVTRNLPDGKADDEPKKRTQYULEDEKTRQLEVL VSPTESSCAQECVETTKKLGNEKQPTNPLEMTKLELIERIAVHIDTISLILQ VAKSISDGDADDEGVPTDQAIRAGLELLKVLSTFHPISFSAETFEESLLACLKMDDE KVAEALAIENKTSKIEDEDPHRISSALLPVLNHRSKSGPRAKVAIHCHALFESSK EFOFQIOTIEPTRLSDPSNLLEHLITPLVTIGHILHLAPDOFAPLKSLVAFIYKDL MNDRLPGKRTKRLWPDDEEVSPEPTVVKTOAIKMYVRLMGKNNHSGSGSTLRILTT ILHSGDLTEOCKISKRPMSRLRLAAGSAIYKLAQECYHEITLDEYQCALAINBE CLQVQAVSRQKILKGLSRRLPLETMAICALCANDPKERRARHROCLVKNINRREY LKQHAVERKELISLLEPYVVPPTJHLAHADPYKVDIEQLKQVCKLMEVLEITMA KNNENSHFEIRKVMENIKOTDAQOPDDAKNNERLYTCDVAMNIINSKSTYSLESP KQVLPAPREFTPQDPKNFSENTKNYLPEPKSFTPTGKPTTYCVAMNIINSKSTYSLESP TKSSEMYNSANSSSNPSPGRILKGRDSEMDHSEHEDYTMSPLRGKKSOKRDS DLVSEIRLEKRGKRTKPTVTOBEETKMDLTKLIOEOKPKGSQSKRRKHTASDDQ QMPBEKRLKEDITLENEDQNSPPKKGKGRPKPLVGGGTPEEFTYKTSKGSKKSG PPAPEPEEEEROSGNTQOKSKOHVRSRAAQRAESPESALISTQSTPQKGRGPP STPSPSPQKRVKLVGRSKQAATKENDSEEDVFGSSPYVDIDIPQEBTEEBEYSTVN VRRASAKRRRR" 235..438
exon	/number=3 439..525
exon	/number=4 526..623
exon	/number=5 624..750
exon	/number=6 751..831
exon	/number=7 832..972
exon	/number=8 973..1088
exon	/number=9 1089..1183
exon	/number=10 1184..1329
exon	/number=11 1330..1481
exon	/number=12 1482..1595
exon	/number=13 1596..1677


```

exon          /number-14
              /number-1726
exon          /number-15
              1727..1866
misc_feature  /number-16
              1866..2046
              /note="matches EST AA687846 from clone IMAGE:1219676"
misc_feature  1866..1941
              /note="matches EST AA687776 from clone IMAGE:1219628"
exon          /number-17
              1867..1982
exon          /number-18
              1983..2088
exon          /number-19
              2089..2249
exon          /number-20
              2250..2373
exon          /number-21
              2374..2532
exon          /number-22
              2533..2601
exon          /number-23
              2602..2738
exon          /number-24
              2739..2862
exon          /number-25
              2863..3067
misc_feature  /number-25
              complement(2912..3302)
              /note="matches EST AA630616 from clone 856164"
exon          /number-26
              3068..3182
exon          /number-27
              3183..3315
misc_feature  /number-27
              3302..3572
              /note="matches EST AA333118"
exon          /number-28
              3572..3751
misc_feature  /number-28
              3751..3818
              /note="matches EST AA282190 from clone IMAGE:704631"
exon          /number-29
              3818..3948
exon          /number-30
              3949..3644
misc_feature  /number-30
              complement(3586..4065)
              /note="matches EST A1655429 from clone IMAGE:2239331"
exon          /number-31
              4065..3750
exon          /number-32
              3751..4190
misc_feature  /number-32
              complement(3864..4286)
              /note="matches EST A1680124 from clone IMAGE:2264521"
misc_feature  4001..4119
              /note="matches EST N83765 from clone KK2745"
misc_feature  4135..4450
              /note="matches EST F12354 from clone c-38h11."
exon          /number-33
              4450..4298
misc_feature  /number-33
              4298..4307
              /note="matches EST T64867 from clone 22264"
exon          /number-34
              4307..4434
misc_feature  /number-34
              4434..4533
              /note="matches EST A4533,4518..4660,4659..4675,4674..4693"
exon          /number-35
              4693..4444
misc_feature  /number-35
              4444..4537
              /note="matches EST A1681422 from clone IMAGE:2272632"
misc_feature  4537..4892
              /note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature  4892..4969
              /note="matches EST A1775866 from clone IMAGE:1877482"
misc_feature  4969..4947
              /note="matches EST AA987361 from clone IMAGE:1603350"

```

```

misc_feature  4538..4943
              /note="matches EST A1299646 from clone IMAGE:1898010"
misc_feature  4943..5003
              /note="matches EST AA447433 from clone 784555"
misc_feature  5003..4571
              /note="matches EST H93424 from clone 220700"
misc_feature  4571..4854
              /note="matches EST A4649,4650..4815"
              /note="matches EST W20070 from clone 306049"
misc_feature  4854..4937
              /note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature  4937..4994
              /note="matches EST AA089775 from clone IMAGE:1269696"
misc_feature  4994..5332
              /note="matches EST AA747568 from clone IMAGE:2329873"
misc_feature  5332..5296
              /note="matches EST AA563884 from clone IMAGE:1013990"
misc_feature  5296..5243
              /note="matches EST D20082 from clone pml283."
misc_feature  5243..4921
              /note="matches EST AA177347 from clone 621810"
misc_feature  4921..5332
              /note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature  5332..4921
              /note="matches EST AA154932 from clone 599194"
misc_feature  4921..4921
              /note="matches EST AW048929 from clone
              UI-M-BH1-amo-C-09-0-UI"
misc_feature  4921..5333
              /note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature  5333..5332
              /note="matches EST A665961 from clone 745478"
misc_feature  5332..5332
              /note="matches EST A1012649 from clone RPLBA02"
misc_feature  5332..4975
              /note="matches EST AA899586 from clone
              UI-R-E0-cy-e-02-0-UI"
misc_feature  4975..5332
              /note="matches EST A1172172 from clone RMBV82"
misc_feature  5332..5332
              /note="matches EST A665961 from clone 745478"
misc_feature  5332..5332
              /note="matches EST A1140790 from clone IMAGE:1566222"
misc_feature  5332..5309
              complement(5021..5309)

Query Match      99.8%; Score 4165.6; DB 9; Length 7444;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4168; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY	421	TATAACATATCTTTGAGTTGGAAGATATGCAATGAAATTTTCCACCCAGCTATACAGAAC	480
Db	547	TATAACATATCTTTGAGTTGGAAGATATGCAATGAAATTTTCCACCCAGCTATACAGAAC	606
QY	481	TTATTTTCAGTTATTAACAAATGGGCACAAATGAGAAAGCCATATGACATGGTAGACCTT	540
Db	607	TTATTTTCAGTTATTAACAAATGGGCACAAATGAGAAAGCCATATGACATGGTAGACCTT	666
QY	541	ATGAGCTCTATTTATTTTGAAGGTAAACAGTGTCTCAGAGCTTTTGGATACGGTTTAA	600
Db	667	ATGAGCTCTATTTATTTTGAAGGTAAACAGTGTCTCAGAGCTTTTGGATACGGTTTAA	726
QY	601	GTAATATCTGGTACTGCTCATTAAGAATTTTAAACAAGACAGATATGATTTGGCAAGGCT	660
Db	727	GTAATATCTGGTACTGCTCATTAAGAATTTTAAACAAGACAGATATGATTTGGCAAGGCT	786
QY	661	TTACGAGAGAGACAGCTCAAGCTATTTGAGCCATATTTTACCATTTTTAAACAGGTT	720
Db	787	TTACTGAGAGAGACAGCTCAAGCTATTTGAGCCATATTTTACCATTTTTAAACAGGTT	846
QY	721	CTGATGCTTTGGGAAAAACATCTATCAGGAGATTTGTACAGACATGCTTTGACTTATTTTG	780
Db	847	CTGATGCTTTGGGAAAAACATCTATCAGGAGATTTGTACAGACATGCTTTGACTTATTTTG	906
QY	781	GAGCTCTTCAATATTTGATAGTACATTTTGTCTGTCTGTCTTTTACCACAGCTTGAATTTAA	840
Db	907	GAGCTCTTCAATATTTGATAGTACATTTTGTCTGTCTGTCTTTTACCACAGCTTGAATTTAA	966
QY	841	TTTAAAGACCAATGATTAATGAGAGAGCGGCTACAACTGTTTAACTACAGGCAAAATGTT	900
Db	967	TTTAAAGACCAATGATTAATGAGAGAGCGGCTACAACTGTTTAACTACAGGCAAAATGTT	1026
QY	901	GGGGCAAAAGATTGAGAAATGGCTTCTCAAAACAAAGCCACTTTTGGCAGTGGCTACTTGGGC	960
Db	1027	GGGGCAAAAGATTGAGAAATGGCTTCTCAAAACAAAGCCACTTTTGGCAGTGGCTACTTGGGC	1086
QY	961	AGGTTTATGATATTCATCTGTACCAAATCCGCTGGAATGTGTGAATTTTGTCTAGCCATTTGT	1020
Db	1087	AGGTTTATGATATTCATCTGTACCAAATCCGCTGGAATGTGTGAATTTTGTCTAGCCATTTGT	1146
QY	1021	CTCATGAAACCAATCCGATTTTAGCAAAAGACTTAACAGATGTCCTTAAAGTAGGTCACAT	1080
Db	1147	CTCATGAAACCAATCCGATTTTAGCAAAAGACTTAACAGATGTCCTTAAAGTAGGTCACAT	1206
QY	1081	GACCCGTGAGAGGATTTATAGACATGATGTATTGTGTCAAAATGTTTACAGCTGCTCAAAAG	1140
Db	1207	GACCCGTGAGAGGATTTATAGACATGATGTATTGTGTGTCAAAATGTTTACAGCTGCTCAAAAG	1266
QY	1141	GATATTTCTTGTGTCATGATGATCTTACTTAAATTTTGTGAGAGAGACAACTTATGACAAA	1200
Db	1267	GATATTTCTTGTGTCATGATGATCTTACTTAAATTTTGTGAGAGAGACAACTTATGACAAA	1326
QY	1201	CGATGGAAGTAGAGCAAAAGAACCATGATGTGGAGCTGGCCCAAAATTTTAAAGAAATATGCT	1260
Db	1327	CGATGGAAGTAGAGCAAAAGAACCATGATGTGGAGCTGGCCCAAAATTTTAAAGAAATATGCT	1386
QY	1261	TTTACAGTCAGAGCTGGAAGAAAGATGCTGCCAAACAGATAGCCATGATCAAAAGCAAAATG	1320
Db	1387	TTTACAGTCAGAGCTGGAAGAAAGATGCTGCCAAACAGATAGCCATGATCAAAAGCAAAATG	1446
QY	1321	CTACATATATATTTCACAAATAGATATGATGATCGACTCTTGTTTGAACGGATCTTTGCT	1380
Db	1447	CTACATATATATTTCACAAATAGATATGATGATGATCGACTCTTGTTTGAACGGATCTTTGCT	1506
QY	1381	CAATACATATGTTCTCCCAATTTTGAATTTGAACAACTACAAAGGATGAATGCTTATATTCTTG	1440
Db	1507	CAATACATATGTTCTCCCAATTTTGAATTTGAACAACTACAAAGGATGAATGCTTATATTCTTG	1566
QY	1441	TATGCGACACATGATTTTAAATGCTGTGTAAGACATTTGATGAAATGTGAAATGTCAAAAT	1500
Db	1567	TATGCGACACATGATTTTAAATGCTGTGTAAGACATTTGATGAAATGTGAAATGTCAAAAT	1626
QY	1501	CTGCTCCGACATCAAGTAAAGATTTGCTGTACTTGAATTAAGCAACCCAAACAGATGCC	1560

Db	1627	CTGCTCGCAGCAATCAAGTAAAGGATTTGGCTTGACTTGATTAAAGCAACCCAAACAGATGCC	1686
QY	1561	AGTGTCAAGGCCATATTTTCAAAGTATGGTTATTAACAAGAAATTTACCGTATCCGTGT	1620
Db	1667	AGTGTCAAGGCCATATTTTCAAAGTATGGTTATTAACAAGAAATTTTACCGTATCCGTGT	1746
QY	1621	AAGGCTCGAGATTTTCATGAAAGAAATTCACACAGGTGTTAGAAAGATGATGAGAAATTAAGA	1680
Db	1747	AAGGCTCGAGATTTTCATGAAAGAAATTCACACAGGTGTTAGAAAGATGATGAGAAATTAAGA	1806
QY	1661	AAGCAGTTAGAAAGTACTTGTTAGTCCAAACATGCTCCGCAAGCAGCGCTGAAGGTGGTG	1740
Db	1807	AAGCAGTTAGAAAGTACTTGTTAGTCCAAACATGCTCCGCAAGCAGCGCTGAAGGTGGTG	1866
QY	1741	CGTGAATAATCAATAGAAAGTTGGGCAACCCCAACAGCTCAAAATCCCTTTCCTGGAATG	1800
Db	1867	CGTGAATAATCAATAGAAAGTTGGGCAACCCCAACAGCTCAAAATCCCTTTCCTGGAATG	1926
QY	1801	ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCATATAGATACCAGATCTATCAGTGT	1860
Db	1927	ATCAAGTTTCTCTTGAGAGAGATAGCACCTGTGCATATAGATACCAGATCTATCAGTGT	1986
QY	1861	CTTATTAACAAGTGAACAAATTCATATAGATGGAACACAGATGTGTAAGATGAGGGTGT	1920
Db	1987	CTTATTAACAAGTGAACAAATTCATATAGATGGAACACAGATGTGTAAGATGAGGGTGT	2046
QY	1921	CCAACGTATCAAGCCATCAGAGCAGGTCCTTAACCTGTTAAGGTACCTCATTTACACAT	1980
Db	2047	CCAACGTATCAAGCCATCAGAGCAGGTCCTTAACCTGTTAAGGTACCTCATTTACACAT	2106
QY	1981	CCCATCTCATTTTCATTTCTGCTGAACAATTTGAATCATTTACTGCTGCTGTAANAATGGAT	2040
Db	2107	CCCATCTCATTTTCATTTCTGCTGAACAATTTGAATCATTTACTGCTGCTGTAANAATGGAT	2166
QY	2041	GATGAAAAAGTAGACAGAAGCTGCACATACAAATTTTCAAAAAACACAGAGACAAATTTGAA	2100
Db	2167	GATGAAAAAGTAGACAGAAGCTGCACATACAAATTTTCAAAAAACACAGAGACAAATTTGAA	2226
QY	2101	GAGATTTTTCACACACATCAGATTCAGACCTTGCTTCCTGTTTATACATCAAAATCTPAAAAA	2160
Db	2227	GAGATTTTTCACACACATCAGATTCAGACCTTGCTTCCTGTTTATACATCAAAATCTPAAAAA	2286
QY	2161	GGACCCCCCGCTCAAGCCAAATATGCCATTCATTGTATCATAGCGATATTTTCTAGTAAA	2220
Db	2287	GGACCCCCCGCTCAAGCCAAATATGCCATTCATTGTATCATAGCGATATTTTCTAGTAAA	2346
QY	2221	GAGACCCAGTTTGGACACAGATATTTTGAGCCTGTGCATAAAGCCTTAGATCCAAAGAACCTG	2280
Db	2347	GAGACCCAGTTTGGACACAGATATTTTGAGCCTGTGCATAAAGCCTTAGATCCAAAGAACCTG	2406
QY	2281	GAACATCTCATATACACACATTTGGTTACTATTTGGTGCATATTTGCTCCCTTGGACCGATCAAA	2340
Db	2407	GAACATCTCATATACACACATTTGGTTACTATTTGGTGCATATTTGCTCCCTTGGACCGATCAAA	2466
QY	2341	TTTGCTGCTCCTTGAAAAATCTTGGGTAGCTACTTTCAATGTGAAAAAGATCTTTCATGAAT	2400
Db	2467	TTTGCTGCTCCTTGAAAAATCTTGGGTAGCTACTTTCAATGTGAAAAAGATCTTTCATGAAT	2526
QY	2401	GATGGGCTTCCAGGGAAAAAGACAACTTAACCTTTGGGTTCCAGATGAAGAAGTATCTCCT	2460
Db	2527	GATGGGCTTCCAGGGAAAAAGACAACTTAACCTTTGGGTTCCAGATGAAGAAGTATCTCCT	2586
QY	2461	GAGCAATATGTCAAAAATTCAGGCTATTTAAATGATGTTGGTGCATGCTGATGATGAAA	2520
Db	2587	GAGCAATATGTCAAAAATTCAGGCTATTTAAATGATGTTGGTGCATGCTGATGATGAAA	2646
QY	2521	AATATATCACAATTAATCAGAACTTCTACCTTAAGATGTGTAACACAAATATTCAGATGT	2580
Db	2647	AATATATCACAATTAATCAGAACTTCTACCTTAAGATGTGTAACACAAATATTCAGATGT	2706
QY	2581	GATGAGACCTTGACAGAACAGGGAAAAATTAGTAACACAGATATGTACGCTTGAGACTT	2640

Db 2707 GATGAGACTGTACAGAACGAGGGAATAATAGTAAACCAAGATATGTCACGCTGAGACTT 2766
QY 2641 GGTGCGGAGGTGTATGTGGAAGCGGACAAAGCCCTGTACATGAATCATCACAA 2700
Db 2767 GGTGCGGAGGTGTATGTGGAAGCGGACAAAGCCCTGTACATGAATCATCACAA 2826
QY 2701 TTGAACAATATCAGCTATGTGCAATAGTATCAACGATGAATGCATCAAGTAAGACA 2760
Db 2827 TTGAACAATATCAGCTATGTGCAATAGTATCAACGATGAATGCATCAAGTAAGACA 2886
QY 2761 GTGTTTCCCAAGAACTTCCAGAGCCCTTCCGTTTACGGCTTCCACTGAGTATATG 2820
Db 2887 GTGTTTCCCAAGAACTTCCAGAGCCCTTCCGTTTACGGCTTCCACTGAGTATATG 2946
QY 2821 GGAATCTGTCGCTTGTGCAAAAGTCTGTAAAGAGAGAAAGCCCTATGCTAGAGCAA 2880
Db 2947 GCAATCTGTCGCTTGTGCAAAAGTCTGTAAAGAGAGAAAGCCCTATGCTAGAGCAA 3006
QY 2881 TGTGTTGCAAAATATAAATGTAAAGCGGAGATCTGAGAGCAGCTGAGCTTACT 2940
Db 3007 TGTGTTGCAAAATATAAATGTAAAGCGGAGATCTGAGAGCAGCTGAGCTTACT 3066
QY 2941 GAAAAATTTATGTCCTTCTTACCAAGATGTTGTTCCATATACATTCACCTTTTGCA 3000
Db 3067 GAAAAATTTATGTCCTTCTTACCAAGATGTTGTTCCATATACATTCACCTTTTGCA 3126
QY 3001 CATGACCCAGATATGTCAAAAGTACAGATTTGAACAATTAACATTTAAAGATGT 3060
Db 3127 CATGACCCAGATATGTCAAAAGTACAGATTTGAACAATTTAAAGATGT 3186
QY 3061 CTTTGGTTTGTCTGGAATAATTTAAATGCTAAATAATGAATTAACAGTACGCTTTATC 3120
Db 3187 CTTTGGTTTGTCTGGAATAATTTAAATGCTAAATAATGAATTAACAGTACGCTTTATC 3246
QY 3121 AGAAAGATGTGTAATAATTTAAACAACAAAGATGCCAAGGACAGATGATGCAAA 3180
Db 3247 AGAAAGATGTGTAATAATTTAAACAACAAAGATGCCAAGGACAGATGATGCAAA 3306
QY 3181 ATGAATGAAAACTGTACACGTGTGTGTTGCCATTAATATCATGTCATCAAGAT 3240
Db 3307 ATGAATGAAAACTGTACACGTGTGTGTTGCCATTAATATCATGTCATCAAGAT 3366
QY 3241 ACTACATACAGTTTGAATCTCTAAGACCCGCTACTACCAAGCTGTTCTTCACTCAA 3300
Db 3367 ACTACATACAGTTTGAATCTCTAAGACCCGCTACTACCAAGCTGTTCTTCACTCAA 3426
QY 3301 CCTGACAGAAATTTCAATACCAAAAAATTAATCTGCTGAAATGAATCATTTTTC 3360
Db 3427 CCTGACAGAAATTTCAATACCAAAAAATTAATCTGCTGAAATGAATCATTTTTC 3486
QY 3361 ACTCTGGAACCTTAAACCAATGTTCTAGAGGCTTTTAAACAGCCTTTTCATCA 3420
Db 3487 ACTCTGGAACCTTAAACCAATGTTCTAGAGGCTTTTAAACAGCCTTTTCATCA 3546
QY 3421 GCAGGCAAGCAATCTCAGACCAATCATCAGCAATGAAGAACTGTAAAGCAATGCAACGAC 3480
Db 3547 GCAGGCAAGCAATCTCAGACCAATCATCAGCAATGAAGAACTGTAAAGCAATGCAACGAC 3606
QY 3481 AGCTCAATTCAGAGCTCTCTGGAAGAAATTAAGGGGAGCTGTATAGTTCTGAATGAT 3540
Db 3607 AGCTCAATTCAGAGCTCTCTGGAAGAAATTAAGGGGAGCTGTATAGTTCTGAATGAT 3666
QY 3541 CACATGTGAATAAGAGATTACACATGTCTTCACTTTCCGGGGAAAAAAGTGACAG 3600
Db 3667 CACATGTGAATAAGAGATTACACATGTCTTCACTTTCCGGGGAAAAAAGTGACAG 3726
QY 3601 AGAGAGCACTGTGATCTTGAAGCTGTGAATTTGAAGAGCACTTGAAGAGCAAGAAAAAGC 3660
Db 3727 AGAGAGCACTGTGATCTTGAAGCTGTGAATTTGAAGAGCACTTGAAGAGCAAGAAAAAGC 3786
QY 3661 CCGGTACAGAAACAGAGAGAAATTTAGTATGATGACTTACTAAGTTGGTACAGAA 3720
Db 3787 CCGGTACAGAAACAGAGAGAAATTTAGTATGATGACTTACTAAGTTGGTACAGAA 3846

QY 3721 CAGAAACCTTAAGGCGATCAGGAGGTGCGAAAAAGAGCCATACGGCTTACAGATCTGAT 3780
Db 3847 CAGAAACCTTAAGGCGATCAGGAGGTGCGAAAAAGAGCCATACGGCTTACAGATCTGAT 3906
QY 3781 GAACAGCACTGCGCTGAGGAAAAAGAGCTCAAAAGAGATATTTAGAAAAATGAATGAA 3840
Db 3907 GAACAGCACTGCGCTGAGGAAAAAGAGCTCAAAAGAGATATTTAGAAAAATGAATGAA 3966
QY 3841 CAGATATGTCGCGCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGT 3900
Db 3967 CAGATATGTCGCGCAAAAAAGGTTAAAGAGCCGACACCAAAACCTTTGGTGAAGT 4026
QY 3901 ACACCAAAAGAGAGCAACATGAAATCTTCTTAAAAAGGAAAGCAAAAAAATTTGGA 3960
Db 4027 ACACCAAAAGAGAGCAACATGAAATCTTCTTAAAAAGGAAAGCAAAAAAATTTGGA 4086
QY 3961 CCTCCAGCACAGAGAGAGAGAAAGAAAGCAAAAGTGAATACGGAACAGAG 4020
Db 4087 CCTCCAGCACAGAGAGAGAGAAAGAAAGCAAAAGTGAATACGGAACAGAG 4146
QY 4021 TCCAAAGCAAAACAGACCGAGTGTCAAGAGAGCACACAGAGACAAATCTCTGAA 4080
Db 4147 TCCAAAGCAAAACAGACCGAGTGTCAAGAGAGCACACAGAGACAAATCTCTGAA 4206
QY 4081 TCTAGTCAATTTGAATCCACAGATCCACACACAGACAGAAAGAGAGACCATCAAA 4140
Db 4207 TCTAGTCAATTTGAATTTCCACAGATCCACACACAGACAGAAAGAGAGACCATCAAA 4266
QY 4141 ACGCCATCACATCACACCAACCAAAAAAATGT 4172
Db 4267 ACGCCATCACATCACACCAACCAAAAAAATGT 4298

Db 4267 ACGCCATCACATCACACCAACCAAAAAAATGT 4298
RESULT 3
HS095825 5253 bp mRNA linear PRI 05-APR-1999
LOCUS
DEFINITION
Human androgen-induced prostate proliferative shutoff associated
protein (AS3) mRNA, complete cds.
VERSION
095825.2 GI:4559409
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
Androgen-induced proliferative shutoff in prostate cancer cells
Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
2 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Early gene expression during androgen-induced inhibition of
proliferation of prostate cancer cells: a new suppressor candidate
on chromosome 13, in the BRCA2-Rb1 locus
J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
JOURNAL
MEDLINE
9929875
PUBMED
10215036
REFERENCE
3 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Direct Submission
JOURNAL
Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
4 (bases 1 to 5253)
Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
Direct Submission
JOURNAL
Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REMARK
Sequence update by submitter
On Apr 5, 1999 this version replaced gl:4539617.
FEATURES
1..5253
/organism="Homo sapiens"

BASE COUNT 1755 a 944 c 1074 g 1480 t
ORIGIN
Query Match 99.28; Score 4141; DB 9; Length 5253;
Best Local Similarity 99.58; Pred. No. 0;
Matches 4153; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 1 ATGGCTCATTTCAAGAGCTAGAGCAATGATGCAAAAATTAACATATCCGCTGGGCTCAAG 60
DB ATGGCTCATTTCAAGAGCTAGAGCAATGATGCAAAAATTAACATATCCGCTGGGCTCAAG 125
QY 61 GAAATATCAGATTAATAATATCTAAAGAGAGATGCTGAGACGATTAAGATGCTTGAAA 120
DB GAAATATCAGATTAATAATATCTAAAGAGAGATGCTGAGACGATTAAGATGCTTGAAA 185
QY 121 ACTTTATGATATGAGACGAGCTCTGAAAGAAAAGAGCTTTATTTAAACCTAGCT 180
DB ACTTTATGATATGAGACGAGCTCTGAAAGAAAAGAGCTTTATTTAAACCTAGCT 245
QY 181 TTACATCTTGCCTTCAGATTTTCTCAAGCATCTGTTAAAGATGCTGCTTACGTGA 240
DB TTACATCTTGCCTTCAGATTTTCTCAAGCATCTGTTAAAGATGCTGCTTACGTGA 305
QY 241 GCCTGCTGCTTGCCTGATATTTTTCAGAGATTTATCTCTGAAGCTCTTACACATCCCT 300
DB GCCTGCTGCTTGCCTGATATTTTTCAGAGATTTATGCTCTGAAGCTCTTACACATCCCT 365
QY 301 GATTAACCTAAAGATATATTTATGTTTATTAACAAGACAGTGAAGGGCTAGAGATACA 360
DB GATTAACCTAAAGATATATTTATGTTTATTAACAAGACAGTGAAGGGCTAGAGATACA 425
QY 361 AAGAGCCCAATTCATAGTATTTTATTTACTGAGAACATTCCTTGGGCTCAAGTCA 420
DB AAGAGCCCAATTCATAGTATTTTATTTTACTGAGAACATTCCTTGGGCTCAAGTCA 485
QY 426 AAGAGCCCAATTCATAGTATTTTATTTGCTTGAAGAACATTCCTTGGGCTCAAGTCA 485
DB AAGAGCCCAATTCATAGTATTTTATTTGCTTGAAGAACATTCCTTGGGCTCAAGTCA 545
QY 481 TATTAACATATGCTTTGAGTTAGAGATTAAGATTAAGATTTTCAACCAAGCTTATACAGAAC 480
DB TATTAACATATGCTTTGAGTTAGAGATTAAGATTAAGATTTTCAACCAAGCTTATACAGAAC 545
QY 481 TATTTTTCAGTTTATTAACAAGATGAGCAATATGACATATGACATATGAGTCACTT 540
DB TATTTTTCAGTTTATTAACAAGATGAGCAATATGACATATGACATATGAGTCACTT 605
QY 541 ATGACCTCATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTTGGATACGGTTTA 600
DB ATGACCTCATTTATTTGTAAGGTGATACAGTGTCTCAGAGCTTTTGGATACGGTTTA 665
QY 601 GTAATATCTGATACCTGCTCATTAAGATTTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
DB GTAATATCTGATACCTGCTCATTAAGATTTTAAACAAGCAAGCATATGATTTGGCAAGGCT 725
QY 661 TTACAGAGAGAGAGAGCTCAAGCTATGAGCAATATATACCACTTTTATATCAGGTT 720
DB TTACAGAGAGAGAGAGCTCAAGCTATGAGCAATATATATACCACTTTTATATCAGGTT 785
QY 721 CTGATGCTTGGGAAACAATCTATCAGAGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
DB CTGATGCTTGGGAAACAATCTATCAGAGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 845
QY 781 GAGCTTACAAATATTTGATAGTCAATTTGCTGCTCTGTTTAAACCCAGCTTGAATTTTAA 840
DB GAGCTTACAAATATTTGATAGTCAATTTGCTGCTCTGTTTAAACCCAGCTTGAATTTTAA 905
QY 841 TTAAGAGCAATGATTAAGAGAGAGGCTCAGCAAGTTGTTAACTACTGCAAAAATGTTT 900
DB TTAAGAGCAATGATTAAGAGAGAGGCTCAGCAAGTTGTTAACTACTGCAAAAATGTTT 965
QY 901 GGGGCAAGAGATTCAGATTTGCTTCAAAAACAGCACTTGGCAGTCTACTTGGGC 960
DB GGGGCAAGAGATTCAGATTTGCTTCAAAAACAGCACTTGGCAGTCTACTTGGGC 1025
QY 961 AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTAATTTGCTAGCCATTCG 1020
DB AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTAATTTGCTAGCCATTCG

DB 1026 AGGTTTAATGATATTCATGTAACAATCCGCTGGAATGTGTAATTTGCTAGCCATTCG 1085
QY 1021 CTCATGACCATTCGATTTAGCAAAAAGCTTAAAGAGATATCTTAAAGAGGTCACAT 1080
DB CTCATGACCATTCGATTTAGCAAAAAGCTTAAAGAGATATCTTAAAGAGGTCACAT 1145
QY 1081 GACCCCTGAGGAAGCTTTAGACATGATGTTATTTGCTCAATAGTTTACACCTGCTAAAAG 1140
DB GACCCCTGAGGAAGCTTTAGACATGATGTTATTTGCTCAATAGTTTACACCTGCTAAAAG 1205
QY 1141 GATATTTCTTGGCTCAATGATCACTTAAATTTTGTGAGAGAGAAACATTAGACAAA 1200
DB GATATTTCTTGGCTCAATGATCACTTAAATTTTGTGAGAGAGAAACATTAGACAAA 1265
QY 1201 CGATGAGAGCTACGCCAAAGAACCATGATGGGCTTGGCCAAATTTTAAAGAAATTCGT 1260
DB CGATGAGAGCTACGCCAAAGAACCATGATGGGCTTGGCCAAATTTTAAAGAAATTCGT 1325
QY 1261 TTACAGTCAAGCTGGAAGAGATGCTGCAAAAACAGATAGCATGATCAAAAGCAAAATG 1320
DB TTACAGTCAAGCTGGAAGAGATGCTGCAAAAACAGATAGCATGATCAAAAGCAAAATG 1385
QY 1321 CTACATATATATTAACAAAATAGTATGATGATGATGATGATGATGATGATGATGAT 1380
DB CTACATATATATTAACAAAATAGTATGATGATGATGATGATGATGATGATGATGAT 1445
QY 1381 CATATCATGCTTCTCACAAATTTAGAACTACAGAAACGATGAAATGCTTATATTCCTG 1440
DB CATATCATGCTTCTCACAAATTTAGAACTACAGAAACGATGAAATGCTTATATTCCTG 1505
QY 1441 TATGCCACACTGATTAATGCTGGAAGCACTTGAAGAAATGGAATGTCACAAAT 1500
DB TATGCCACACTGATTAATGCTGGAAGCACTTGAAGAAATGGAATGTCACAAAT 1565
QY 1501 CTGCTCCGACATCAAGTAAGATTTGCTTGAAGCTTGAATTAAGCAACCAACAGATGCC 1560
DB CTGCTCCGACATCAAGTAAGATTTGCTTGAAGCTTGAATTAAGCAACCAACAGATGCC 1625
QY 1561 AGTGTCAAGGCTATTTTCAAAAGATGATGATTTATCAAGAAATTTACTGATCCTGCT 1620
DB AGTGTCAAGGCTATTTTCAAAAGATGATGATTTATCAAGAAATTTACTGATCCTGCT 1685
QY 1621 AAGGCTCAGAGATTTTCAAGAAATTTCAACAGGCTTGAAGATGATGAGAAATTAAGA 1680
DB AAGGCTCAGAGATTTTCAAGAAATTTCAACAGGCTTGAAGATGATGAGAAATTAAGA 1745
QY 1681 AAGCAATTAAGAGTACTTGTAGTCCACATGCTCTGCAAGCAGGCTGAAGGTTGTGTG 1740
DB AAGCAATTAAGAGTACTTGTAGTCCACATGCTCTGCAAGCAGGCTGAAGGTTGTGTG 1805
QY 1741 CGTGAATTAAGTAAGAGTTGGGCAACCCCAACAGCCCTTCAAAATCCTTCTGGAAAG 1800
DB CGTGAATTAAGTAAGAGTTGGGCAACCCCAACAGCCCTTCAAAATCCTTCTGGAAAG 1865
QY 1801 ATCAAGATTTCTTGGAGAGATAGCACCTGTGACATATGATACCGAAATCTATCAGTGT 1860
DB ATCAAGATTTCTTGGAGAGATAGCACCTGTGACATATGATACCGAAATCTATCAGTGT 1925
QY 1861 CTATTTAAACAAGTAACAATCAATAGATGAGACAGCATGATGAGAGTGGTGT 1920
DB CTATTTAAACAAGTAACAATCAATAGATGAGACAGCATGATGAGAGTGGTGT 1985
QY 1921 CCAACTGATCAAGCCATCAGACAGGCTGTGAACCTGTTAAAGTACTCATTATTCACAT 1980
DB CCAACTGATCAAGCCATCAGACAGGCTGTGAACCTGTTAAAGTACTCATTATTCACAT 2045
QY 1981 CCCATCTCATTTCTGCTGAACATTTTGAATATTTACTGCTTGTGAAAATGAT 2040
DB CCCATCTCATTTCTGCTGAACATTTTGAATATTTACTGCTTGTGAAAATGAT 2105
QY 2041 GATGAAAAAGTAGAGAGAGCTGCACTACAAATTTTCAAAAACACAGAGCAAAATTTGAA 2100
DB GATGAAAAAGTAGAGAGAGCTGCACTACAAATTTTCAAAAACACAGAGCAAAATTTGAA 2165

QY 2101 GAGATTTTCCACATCAGATCAGCCCTTGCCTGTTTTTACATCACAATCTTAAAAA 2160
 |||||
 Db 2166 GAGATTTTCCACATCAGATCAGCCCTTGCCTGTTTTTACATCACAATCTTAAAAA 2225
 |||||
 QY 2161 GGACCCCCCGTCAAGCCAAATATGCCATTTATGATCATGCGATTTTTTCTAGTAAA 2220
 |||||
 Db 2226 GGACCCCCCGTCAAGCCAAATATGCCATTTATGATCATGCGATTTTTTCTAGTAAA 2285
 |||||
 QY 2221 GAGACCCAGTTTGCACAGATATTTGAGCCCTGCGATTAAGAGCCATGACCAAGCAAGCTG 2280
 |||||
 Db 2286 GAGCCCGAGTTTGCACAGATATTTGAGCCCTGCGATTAAGAGCCATGACCAAGCAAGCTG 2345
 |||||
 QY 2281 GAACATCTCATACACACATGTTACTATATGTCATATTTGCTCTCCACCTGATCAA 2340
 |||||
 Db 2346 GAACATCTCATACACACATGTTACTATATGTCATATTTGCTCTCCACCTGATCAA 2405
 |||||
 QY 2341 TTTGCTGCTCTGGAATCTTTGGTACTACTTTGATTTGGAAGATCTTCTCATGAT 2400
 |||||
 Db 2406 TTTGCTGCTCTGGAATCTTTGGTACTACTTTGATTTGGAAGATCTTCTCATGAT 2465
 |||||
 QY 2401 GATCGGCTTCCAGGAAAAAACAACCTTAACCTTTGGTCCAGATGAAGAATCTGCT 2460
 |||||
 Db 2466 GATCGGCTTCCAGGAAAAAACAACCTTAACCTTTGGTCCAGATGAAGAATCTGCT 2525
 |||||
 QY 2461 GAGACATGTCAAAATTCAGGCTATTAATAATGATGTTGATGGCTACTTGGAAATGAAA 2520
 |||||
 Db 2526 GAGACATGTCAAAATTCAGGCTATTAATAATGATGTTGATGGCTACTTGGAAATGAAA 2585
 |||||
 QY 2521 AATATCATGATTAATCAGAACTTCTACCTTAAGATGCTTAACAACAATTTGATAGT 2580
 |||||
 Db 2586 AATATCATGATTAATCAGAACTTCTACCTTAAGATGCTTAACAACAATTTGATAGT 2645
 |||||
 QY 2581 GATGAGACTTGACAGAAACAGGGGAAAAATTAGTAACAGATATGTCAGCTGAGACTT 2640
 |||||
 Db 2646 GATGAGACTTGACAGAAACAGGGGAAAAATTAGTAACAGATATGTCAGCTGAGACTT 2705
 |||||
 QY 2641 GCTGCTGGAGTGTCTATTTGTGAAGTGGCACAAGAACCTTGTACATGAATCTACACA 2700
 |||||
 Db 2706 GCTGCTGGAGTGTCTATTTGTGAAGTGGCACAAGAACCTTGTACATGAATCTACACA 2765
 |||||
 QY 2701 TTAGACAATATACGCTATGTCATTTACTATCAACGATGATGATCAAGTAAGACAA 2760
 |||||
 Db 2766 TTAGACAATATACGCTATGTCATTTACTATCAACGATGATGATCAAGTAAGACAA 2825
 |||||
 QY 2761 GTGTTTGGCCAGAACTTCCAAAGGCTTTCCGTTTACGCGCTCCACTTGAGTATATG 2820
 |||||
 Db 2826 GTGTTTGGCCAGAACTTCCAAAGGCTTTCCGTTTACGCGCTCCACTTGAGTATATG 2885
 |||||
 QY 2821 GCAATCTGTGCTTTGTGCAAAAAGATCTGTAAAGAGAGAGAGAGCTCATGCTAGGCAA 2880
 |||||
 Db 2886 GCAATCTGTGCTTTGTGCAAAAAGATCTGTAAAGAGAGAGAGAGCTCATGCTAGGCAA 2945
 |||||
 QY 2881 TGTGTTGGTGAATAATATTAATGTAAGGGGAGATCTGGAAGCAGATGCAAGCTTATG 2940
 |||||
 Db 2946 TGTGTTGGTGAATAATATTAATGTAAGGGGAGATCTGGAAGCAGATGCAAGCTTATG 3005
 |||||
 QY 2941 GAAAAATATATGCTCTCTACAGAGATGTTGTCATATACATTCACCTTTGGCA 3000
 |||||
 Db 3006 GAAAAATATATGCTCTCTACAGAGATGTTGTCATATACATTCACCTTTGGCA 3065
 |||||
 QY 3001 CATGACCCAGATTTATGTCAAAAGTACAGATATTTGAACAACCTTAAGATGTTAAAGATGT 3060
 |||||
 Db 3066 CATGACCCAGATTTATGTCAAAAGTACAGATATTTGAACAACCTTAAGATGTTAAAGATGT 3125
 |||||
 QY 3061 CTTTGGTTTGTGGAATAATTAATGCTTAATAATGAATAATACAGTACAGCTTTATC 3120
 |||||
 Db 3126 CTTTGGTTTGTGGAATAATTAATGCTTAATAATGAATAATACAGTACAGCTTTATC 3185
 |||||
 QY 3121 AGAAAGATGTAGAAAATATTAACAAACAAAGATGCCAAGGACGACGATGATGCAAAA 3180
 |||||
 Db 3186 AGAAAGATGTAGAAAATATTAACAAACAAAGATGCCAAGGACGACGATGATGCAAAA 3245
 |||||

QY 3181 ATGAATGAAAACTGTACACTGTGTGATGTTGGCATGAATATCATCATGTCAAAGAT 3240
 |||||
 Db 3246 ATGAATGAAAACTGTACACTGTGTGATGTTGGCATGAATATCATCATGTCAAAGAT 3305
 |||||
 QY 3241 ACTACATCAGTTTGGAAATCTCCTAAGACCCCGTACTACAGCTCGTTTCTTACATCAA 3300
 |||||
 Db 3306 ACTACATCAGTTTGGAAATCTCCTAAGACCCCGTACTACAGCTCGTTTCTTACATCAA 3365
 |||||
 QY 3301 CCTGCAAGAAATTTCAATACACAAAAAATATATCGCTCCCTGAAATGAATCAATTTTTC 3360
 |||||
 Db 3366 CCTGCAAGAAATTTCAATACACAAAAAATATATCGCTCCCTGAAATGAATCAATTTTTC 3425
 |||||
 QY 3421 GCAGCAAGCAATCTCAGCAACCAATTCACAGANTGGAACCTTAAGCAATGCAAGCAGC 3480
 |||||
 Db 3486 GCAGCAAGCAATCTCAGCAACCAATTCACAGANTGGAACCTTAAGCAATGCAAGCAGC 3545
 |||||
 QY 3481 AGCTCAATCCAGAGCTCTCTGAAAGATTAAGGGGAGGCTTGATAGTTCTGAATGGAT 3540
 |||||
 Db 3546 AGCTCAATCCAGAGCTCTCTGAAAGATTAAGGGGAGGCTTGATAGTTCTGAATGGAT 3605
 |||||
 QY 3541 CACAGTGAATAATGAAGATTTACACAATGCTTCACCTTTCCGGGGAAAAAAGTGACAA 3600
 |||||
 Db 3606 CACAGTGAATAATGAAGATTTACACAATGCTTCACCTTTCCGGGGAAAAAAGTGACAA 3665
 |||||
 QY 3601 AGAGCGACTCGTGAATCTTGAAGCTGTAATTTGGAAGACCTTAGAGCGGAGAAAAAAGC 3660
 |||||
 Db 3666 AGAGCGACTCGTGAATCTTGAAGCTGTAATTTGGAAGACCTTAGAGCGGAGAAAAAAGC 3725
 |||||
 QY 3661 CCCGTACACAGAACAGAGAGAAATTAAGTATGATGACTTGAATTTGTTGTAACAGAA 3720
 |||||
 Db 3726 CCCGTACACAGAACAGAGAGAAATTAAGTATGATGACTTGAATTTGTTGTAACAGAA 3785
 |||||
 QY 3721 CAGAACTTAAGAGAGTACAGAGAGTGGAAAAAGAGCCCTAGCGCTTCGATCTGAT 3780
 |||||
 Db 3786 CAGAACTTAAGAGAGTACAGAGAGTGGAAAAAGAGCCCTAGCGCTTCGATCTGAT 3845
 |||||
 QY 3781 GAACAGCAGTGGCTGAGGAAAAAGAGGCTCAAGAAGATATTAAGAAATGAAGATGAA 3840
 |||||
 Db 3846 GAACAGCAGTGGCTGAGGAAAAAGAGGCTCAAGAAGATATTAAGAAATGAAGATGAA 3905
 |||||
 QY 3841 CAGATAGTCCGCCAAAAAAGGTTAAAGAGAGCCGACCCAAAAAAGCTTTGTTGAGAGT 3900
 |||||
 Db 3906 CAGATAGTCCGCCAAAAAAGGTTAAAGAGAGCCGACCCAAAAAAGCTTTGTTGAGAGT 3965
 |||||
 QY 3901 ACACCAAAAGAGAGCCCAACATGAAACTTTAATAAAGAGAGCAAAAAAATCTGGA 3960
 |||||
 Db 3966 ACACCAAAAGAGAGCCCAACATGAAACTTTAATAAAGAGAGCAAAAAAATCTGGA 4025
 |||||
 QY 3961 CCTCAGCAGCAG 4020
 |||||
 Db 4026 CCTCAGCAGCAG 4085
 |||||
 QY 4021 TCCAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGACAGCAGAGAGAGAGAGAGAGAGAG 4080
 |||||
 Db 4086 TCCAAAAGCAAAAGCAGCAGAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4145
 |||||
 QY 4081 TCTAGTGAATTTGAATCCACACAGTCCACACCACAGAAAGAGAGAGAGAGAGAGAGAG 4140
 |||||
 Db 4146 TCTAGTGAATTTGAATCCACACAGTCCACACCACAGAAAGAGAGAGAGAGAGAGAGAG 4205
 |||||
 QY 4141 ACGCATCAGCATCAGCAACCAAAAAAATGTG 4173
 |||||
 Db 4206 ACGCATCAGCATCAGCAACCAAAAAAATGTG 4238
 |||||

RESULT 4
 AY102267 5317 bp mRNA linear ROD 01-JUN-2003
 LOCUS Mus musculus androgen-induced prostate proliferative shutoff
 DEFINITION

associated protein AS3 mRNA, complete cds.
AY102267
VERSION
AY102267.1 GI:31321922
KEYWORDS
SOURCE
ORGANISM
Mus musculus
(house mouse)
Eukaryota; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
1 (bases 1 to 5317)
Geck, P., Maffini, M., Sonnenschein, C. and Soto, A.M.
The AS3 proliferative arrest gene has an ancient eukaryotic
heritage and shows highly conserved functional domains in mice
Proc. Ann. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
2 (bases 1 to 5317)
Geck, P., Maffini, M., Liang, S.L., Sonnenschein, C. and Soto, A.M.
Extreme domain conservation and expression of the mouse AS3
proliferative arrest protein
Unpublished
JOURNAL
TITLE
3 (bases 1 to 5317)
Geck, P., Maffini, M., Liang, S.L., Sonnenschein, C. and Soto, A.M.
Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
JOURNAL
AUTHORS
TITLE
Location/Qualifiers
1. 5317
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="5"
/map="B4CM; G2; between BRCA2 and KL"
/sex="male"
/tissue_type="brain"
147. 4487
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein AS3"
/protein_id="AA052216.1"
/db_xref="GI:31321923"
/translation="MAHSKRTINDGKITYPGCVKEISDKISKSEENVRLKMYKTFMD
MDQSEEEKLNLALHLPADFLKHPKDVRLVACCLADIFRIYAEAYTSPDK
LKDIMEFTROLKLEDRKSPQNFRTYLENIAMWKSYNIEFELEDSEIFTOLERT
LESVINGNOCVHMHVDMSSIIPEGDTVOELDLVNLVPHAKMLNKOADLA
KALKRTAOAIEPYITNFENOVLMIGKTSISDSEHVPYITLIEVINISHLISVLPD
LEFKSNDNERLOYVILLAKMFGAKSELSQNKPLMOCTIGRNDIHYVIRLECY
KRASHCLMHPDLAKDLETLKVRSHDEPRLHDVTVSIVPAVKDILLVDHLNF
VREKTLDRWRVRKEMMGLQIVRKYSLSQAAGDAAOISWAKDKLLIVDHLNF
DLILVERIFAQVMPVHNLEETTERMKCLYLYVATLDLNAKALNEMKCCNLRHQVAD
LIDLKORPTDASVKAIFSKVAVITRNLDPEKADPFMKFTQVLEDDERIKROLEAD
VSPSCSKOAEQVREITRKLGPNROPTPPLFEMTKELERLAPVHIDESIILIKO
VVKSIDGTADDEGVPTDOAIRAGLETLKVSPTHPISFHEAETPESTLACKMDE
KYAEAAIOIFKRTGSKIEDPPIHSALLPVLHHKSKPPPOAKAIHCHTAITSK
ETQFAOIFEPRLKSLDPSKLEHLITPLVTIGHTIALAPQFAPLKSLVATFIVDLT
EHLDPGRKTKRLWVDEVSPEYVVKIOAIKMYRWMLGKMNHNSKSTSLRLVLT
ILHSDGLEDOCKISKPDMSRLRLAAGSAIVKLAQEPCHYEITTEQOYLALAINDE
CYOVROFAOKLHKGISRLRLPLEYMAICALADPKVRKRAHAROCUKNTVIREY
LKOHASVEKILSLPEYVYPTIHLADPVRVYODIEOLKDKCEKLVLELTAKMDE
KNNNSHAFIRKVENIKOTKDAQOPDDTKNMEKLYTVDAVNMNIMSSTYSLESP
KDPVLPAREFTOPDKNFSNTKNTLPEPKMSFTTPEKPTANVLAANKLSAGKQSO
TKSSHMETVSNASSNSPSPGRIGRLDSEMDHSENDYMSPLDPKSKDREDDP
DLSELEKPRSKKAPVTDPEERKIGMDLTKLVEQOKPKSGRGRKGRASDSDEQM
PEERKHKEELENEDEONSPKPKGRGRPKPLGGSTKEETPMKTSKKKKKLVPV
VYDDDEEERIOIGNTEHKSQKQHTSKRAOORASPEFSVNESQSTQKGRGRPK
APSPQPKKIKVGRSKOYATREKENDSEEMDVLAQSSPVSDTTOBGAEEDISGVN
RRKSKRRER"

BASE COUNT 1676 a 1100 c 1199 g 1342 t
ORIGIN
Query Match 83.2% Score 3471.8; DB 10; Length 5317;
Best Local Similarity 89.7%; Pred. No. 0;
Matches 3741; Conservative 0; Mismatches 422; Indels 6; Gaps 1;
QY 1 ATGGCTCATTTCAAGACTAGACCAATGAGAAAATTACATATCCGCTGGGCTCAAG 60

|||||
147 ATGGCTCATTTCAAGAGCAGACGACGATGGGAAAATTACTTACCCTCTGGAGTCAAG 206
QY 61 GAAATATCAGATTAATAATATCTAAGAGAGATGTCAGACGATTAAGATGTTGTGAAA 120
|||||
207 GAAATCAGATTAATAATCTCTAAGAGAGATGTCAGAGCGGTTAAGATGTTGTGAAA 266
QY 121 ACTTTATGATATGACACGACGCTGTGAAGAGAAAAGAGGTTTATTTAAACCTAGCT 180
|||||
267 ACTTTCATGACATGAGACGACGACGCTGTGAAGAGAAAAGAACTTTATCTAAACCTAGCT 326
QY 181 TTACATCTTGCTTCAGATTTTTTTTCTCAAGCATCTGTAAGAGATGTCGCTACGTA 240
|||||
327 TTACATCTTGCTTCAGATCTTCTCCCAAGCATCTGTAAGAGATGTCGCTACGTA 386
QY 241 GCCCTGCTGCTGCTGATATTTTTCAGAGATTTATGCTCTGAAAGCTCTTACACATCCCT 300
|||||
387 GCTTCTGCTGCTGCTGATATTTTTCAGAGATTTATGCTCTGAGAGCTCTTACAGTCTCCC 446
QY 301 GATTAACATGAGATATATTTATGTTTATTAACAGACAGTGTGAAGGGCTAGAGATACA 360
|||||
447 GATTAACATGAGATATATTTATGTTTATTAACAGACAGTGTGAAGGGCTAGAGATACA 506
QY 361 AAGAGCCCAAAATTCATAGATATTTTATTTACTTGAGACATTTGCTTGAGTCAAGTCA 420
|||||
507 AAGAGCCCTCAGTTTATATAGTATTTTATTTACTTGAGACATTTGAGTCAAGTCA 566
QY 421 TATTAACATATGCTTTGAGTTAGAGATGATGAAATTTTTCACCCAGCTATACAGATCA 480
|||||
567 TATTAACATATGCTTTGAGTTAGAGATGATGAAATTTTTCACCCAGCTATACAGATCA 626
QY 481 TTATTTTCAGTATATAACATATGCGCCCATTCAGAAAGTCCATATGACATGATGACCTT 540
|||||
627 TTATTTTCAGTATATAACATATGCGCCCATTCAGAAAGTCCATATGACATGATGACCTT 686
QY 541 ATGAGCTATATTTTGTGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
|||||
687 ATGAGCTATATTTTGTGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
QY 747 GTAATATCTGTCACGTCGCCCATTAAGAACTTAACAAAGCAAGCATATGATTTGGCAAGGCT 806
QY 601 GTAATATCTGTCACGTCGCATTAAGAACTTAACAAAGCAAGCATATGATTTGGCAAGGCT 660
|||||
807 TTGCTGAAGAGAGATGCTCAAGCTATGACCATATATTAACCAATTTTAAACAGGTT 866
QY 721 CTGATGCTTTGGGAAAACATCTATACAGCGATTTGTGACAGCATGCTTTGACTTAATTTTG 780
|||||
867 CTGATGCTTTGGGAAAACATCTATACAGTATTTGTGACAGCATGCTTTGACTTAATTTTG 926
QY 781 GAGCTCTACATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
|||||
927 GAGCTCTACATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 986
QY 841 TTAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
|||||
987 TTAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1046
QY 901 GGGGCAAAAGATTCAGAAATTTGGCTTCTCAAAACAGCCACTTTGGGAGTCTACTTTGGGC 960
|||||
1047 GGGGCAAAAGATTCAGAAATTTGGCTTCTCAAAACAGCCACTTTGGGAGTCTACTTTGGGC 1106
QY 961 AGTTTATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
|||||
1107 AGTTTATGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1166
QY 1021 CTCATGACCATCTGATTTAGCAAAAGACTTAACAGAGATGATTTAAAGTGAAGTACAT 1080
|||||
1167 CTCATGACCATCTGATTTAGCAAAAGACTTAACAGAGATGATTTAAAGTGAAGTACAT 1226
QY 1081 GACCCGTGAGAGATCTATTTAGACATGATGTTTATGTTGTCATATGATTTACACCTCTTAAAG 1140
|||||

Dd 1227 GATCCTGAGAGACTATTAGACATGATGTTATGTGCTATATAGTAAACGCTGCTAAAAAG 1286
Qy 1141 GATATTCCTTGTGCAATGATGACTTACTTAATTTTGTGAGAGAGAGAACATTTAGACAAA 1200
Dd 1287 GATATTCCTTGTGCAATGATGACTTACTTAATTTTGTGAGAGAGAGAACATTTAGACAAA 1346
Qy 1201 CGATGAGAGATGACCAAGAAAGCCATGATGGGACTTGGCCAAATTTATAGAAATATGCT 1260
Dd 1347 CGGTGGAGAGTACCTAAAGAACCATGATGGGACTGGCCAGATTTTATAAAGAAATACTCT 1406
Qy 1261 TTACAGTACGACGCTGGAAAGATGCTGCAAAACAGATAGATGATGATCAAGACAAATG 1320
Dd 1407 TTACAGTACGACGCTGGAAAGATGCTGCAAAACAGATATCTGGGCTCAAGGCAAGCTG 1466
Qy 1321 CTACATATATATTTTCAAAATAGTATGATGATGACTTGTGTAAGCGGATCTTGTCT 1380
Dd 1467 CTGATATATATCTTCAAGAACAGCATGACGCGGCTACTTGTGAGCGGATCTTGTCT 1526
Qy 1381 CAATGACATGTTCTCTCACATTTTGAACCTACAGACGGATGAATGCTTATATCTTG 1440
Dd 1527 CAGTACATGTCCTCTCACAAATTTGAAACGACGAGACGATGAAGTCTGTATTTACTTG 1586
Qy 1441 TATGCAACATGATTTAAATGCTGTGAAGCATTTGAATGAAATGCTAAATGCTAAAT 1500
Dd 1587 TACGCTACACTGGAATGATGCTGTGAAGCATTTGAATGAAATGCTAAATGCTAAAT 1646
Qy 1501 CTGCTCGACATCAAGTAAAGATTTGCTTACTTGTATTAAGCAACCCAAACAGATGCC 1560
Dd 1647 CTACTCGACATCAAGTAAAGATTTTACTTATTTAATTAACAACCAAAACAGATGCC 1706
Qy 1561 AGTGTCAAGGCCATATTTTCAAAAGTATGTTATTCACAGAAATTTACCTGATCTGCT 1620
Dd 1707 AGTGTCAAGGCCATATTTTCAAAAGTATGTTATTCACAGAAATTTCCAGATCTGCT 1766
Qy 1521 AAGCTCAGATTTCAAGAAATTCACAGAGGTGAGAAAGATGAGAAAAATAGA 1680
Dd 1767 AAGGCCCAAGCTTCATGAGAAAGTTTCAACAGAGGTGAGAAAGATGAGAAAAATAGA 1826
Qy 1681 AAGCAGTTAGAACTTGTATGTCACACATGCTCTGTCACAGACGCTGAAGGTGTGTG 1740
Dd 1827 AAACAGGTGAAGCCCTGTATGTCACATGTTCTGTCACAGACGCTGAAGGCTGTGTG 1886
Qy 1741 CGTGAATTAATTAAGAGTTGGGCAACCCCAAGACCTTACAAATCTTTCTGTGAAATG 1800
Dd 1887 CGTGAATTAATTAAGAGTTGGGCAACCCCAAGACCTTACAAATCTTTCTGTGAAATG 1946
Qy 1801 ATCAAGTTCTCTGAGAGAGATGACACCTGTCACATAGATAGCGAATCTATAGTCT 1860
Dd 1947 ATCAAGTTCTCTGAGAGAGATGACACCTGTCACATAGACACTGAATCCATCAGTCT 2006
Qy 1861 CTATTTAAACAAGTGAACAAATCAATAGATGAACAGCAGATGATGAAGTGAAGGTGT 1920
Dd 2007 CTATTTAAACAAGTGAACAAATCAATAGATGAACAGCAGATGATGAAGTGAAGGTGT 2066
Qy 1921 CCAACTGATCAAGCATGACGACAGGTCTTGAAGTCTTAAAGTACTCTCAATTAACAT 1980
Dd 2067 CCAACTGATCAAGCATGACGACAGGTCTTGAAGTCTTGAAGTCTTCAATTAACGAC 2126
Qy 1981 CCCATCATATTTCTGCTGGAACAAATTTGAATCATTTAGGCTTGTGCTAAATATGAT 2040
Dd 2127 CCTATTCCTTTCTGCTGCTGAGAGCTTGTGATGCTCTTGTGCTGCTGCTGAAGGTGAT 2186
Qy 2041 GATGAAAAAGTGAAGAGTGAACAAATTTTCAAAAAACAGAGAGAGCAAAATTTGAA 2100
Dd 2187 GATGAAAAAGTGAAGAGTGAACAAATTTTCAAAAAACAGAGAGAGCAAAATTTGAA 2246
Qy 2101 GAGATTTTCCACATGATGATGACGCTTGTCTGCTTTTACATCACAATCTTAAAAA 2160
Dd 2247 GAGATTTTCCCTCACATGATGATGACGCTTGTCTGCTTACATCACAATCTTAAAAA 2306
Qy 2161 GGAGCCCCCGCTCAAGCAAAATATGCAATCATTTGATGATGATGATTTTTCAGTAAA 2220
Dd 2307 GGACCACTCGGCGAGGCCAAATAGCCATTTGTATTCATGCCAATTTTTCAGTAAA 2366

Qy 2221 GAGACCCAGTTTGGACAGATATTTGAGCCTCTGATAGAGACCTAGATGCCAAGCACTG 2280
Dd 2367 GAAACCCAGTTTGGACAGATATTTGAGCCTCTGATAGAGACCTAGATGCCAAGCACTG 2426
Qy 2281 GAACATCTCAATTAACACATTTGCTTACTTATTTGCTATTTGCTCTCTTGCACCTGATCA 2340
Dd 2427 GAGCATCTTATTAACCCCTCTGCTCAATTTGCTATTTGCTCTCTTGCACCTGATCA 2486
Qy 2341 TTGCTGCTCTTGGAAATCTTGGGTGCTCTTCAATTTGAAAGATCTTCTCATGAT 2400
Dd 2487 TTGCTGCTCTTGGAAATCTTGGGTGCTCTTCAATTTGAAAGATCTTCTCATGAT 2546
Qy 2401 GATGCGCTTCCAGGGAAGAAAGACAACTTAACTTTGGTTCAGATGAAAGATATCTCT 2460
Dd 2547 GACGCGCTTCCAGGGAAGAAAGACAACTAAGCTTTGGGTTCAGATGAAAGATATCTCA 2606
Qy 2461 GAGACATGTCACAAATTTCAAGGCTATTTAAATGATGTTGCTGATGGCTACTTGAATGAA 2520
Dd 2607 GAGACAAATGTCACAAATTTCAAGGCTATTTAAATGATGTTGCTGATGGCTACTTGAATGAA 2666
Qy 2521 AATATCAAGTAAATCAAGAACTTCTACCTTAAGATGCTTAAACAACAATTTGATAGT 2580
Dd 2667 AATATCAAGTAAATCAAGAACTTCTACCTTAAGATGCTTAAACAACAATTTGATAGT 2726
Qy 2581 GATGAGACTTGAAGAGAGAGGGAATTTAGTAAACAGATATGTCACGCTGAGACTT 2640
Dd 2727 GATGAGACTTGAAGAGAGAGGGAATTTAGTAAACAGATATGTCACGCTGAGACTT 2786
Qy 2641 GCTGCTGGAGTCTTATTTGTGAAGTGGCACAGAACCTTTTACATGAATATATCA 2700
Dd 2787 GCTGCTGGAGTCTTATTTGTGAAGTGGCACAGAACCTTTTACATGAATATATCA 2846
Qy 2701 TTAGCAATATTCAGCTATGCTATGCTATTCACAGTGAATGCTATCAAGTAAAGCA 2760
Dd 2847 CTGAGAGCTACCAAGCTTGTGCTATTCACATGATGATGATGATGATGATGATGATGAT 2906
Qy 2761 GTGTTGGCCAGAAACCTTACAAAGGCTTTCCGTTTAAAGGCTTCCACTTGAATATG 2820
Dd 2907 GTGTTGGCCAGAAACCTTACAAAGGCTTTCCGTTTAAAGGCTTCCACTTGAATATG 2966
Qy 2821 GCAATCTGCTGCTTTTGTGCAAAAGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Dd 2967 GCAATCTGCTGCTTTTGTGCAAAAGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 3026
Qy 2881 TGTGTTGTAAGAAATATTAATGTAAGGCGGAGATATGTAAGAGAGAGAGAGAGAGAG 2940
Dd 3027 TGTGTTGTAAGAAATATTAATGTAAGGCGGAGATATGTAAGAGAGAGAGAGAGAGAG 3086
Qy 2941 GAAAAATATTTGCTCTTCTACAGAGATATGTTCCATTTACAAATTCACCTTTTGCA 3000
Dd 3087 GAAAAATATTTGCTCTTCTACAGAGATATGTTCCATTTACAAATTCACCTTTTGCA 3146
Qy 3001 CATGACCAAGTATATGCAAGATACAGATATTTGAACACTTTAAAGATTTAAAGATGT 3060
Dd 3147 CATGACCAAGTATATGCAAGATACAGATATTTGAACACTTTAAAGATTTAAAGATGT 3206
Qy 3061 CTTGTTGTTGTTGGAATATTAATGTTAGGCTTAAATGAAATTAACAGTCAAGCTTTATC 3120
Dd 3207 CTTGTTGTTGTTGGAATATTAATGTTAGGCTTAAATGAAATTAACAGTCAAGCTTTATC 3266
Qy 3121 AGAAAGATGTAAGAAATATTTAAACAAACAAAGATGCGCAAGGACAGATGATGCAAA 3180
Dd 3267 AGAAAGATGTAAGAAATATTTAAACAAACAAAGATGCGCAAGGACAGATGATGCAAA 3326
Qy 3181 ATGAATGAAAAACGTACACTGTGTGATGTTGCGCATGAATATCATCATCTCAAGAGT 3240
Dd 3327 ATGAATGAAAAATTTGACACCGTGTGATGTTGCGCATGAATATCATCATCTCAAGAGC 3386
Qy 3241 ACTCATTCAGTTTGAATCTCTTAAAGACCGGACTACAGGCTGTTTTCACATCA 3300
Dd 3387 ACCACGTACAGCTTGAAGTCTCTTAAGAGACCCCGTGTGCGACGCTGTTTTCACATCA 3446


```

QY 3301 CCTGACAAAGATTTTCAGTAAACACAAATAATCTCTCTGTAATGATCATTTTTC 3360
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3447 CCTGACAAAGATTTTATGTAACACCAAAATTTACTCTCTCCAGAAATGAAATCTTTTC 3506
QY 3361 ACTCTGAAAACTTAAACCAACCAATGTTCTAGAGACTGTTAACAGCCACTTTTCATCA 3420
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3507 ACTCTGAAAACTTAAACCAACCAATGTTCTAGAGACTGTTAACAGCCACTTTTCATCA 3566
QY 3421 GGAGCAAGCAATCTCAGACCAATCATACCAATGGAACCTGTAAGCAATGCAAGCAC 3480
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3567 GGAGCAAGCAATCTCAGACCAATCATACCAATGGAACCTGTAAGCAATGCAAGCAC 3626
QY 3481 AGCTCAAAATCCAGCTCTCTGGAAGATAAAGGGAGGCTGTAGTACTGTAATGAT 3540
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3627 AGCTCAAAATCCAGCTCTCTGGAAGATAAAGGGAGGCTGTAGTACTGTAATGAT 3686
QY 3541 CACAGTGAATGAAGATTACCAATGTTCTTCACTTTCCGGGGAAAAAAGTGACAG 3600
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3687 CACAGTGAATGAAGATTACCAATGTTCTTCACTTTCCGGGGAAAAAAGTGACAG 3746
QY 3601 AGAGCACTCTGATCTTCTAGTCTGTAATGGAAGCAAGCTTACAGGCAAGAAAAAG 3660
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3747 AGAGCACTCTGATCTTCTAGTCTGTAATGGAAGCAAGCTTACAGGCAAGAAAAAG 3800
QY 3661 CCCGTCACAGAACAGGAGAGAAATTAGTATGATGACTTACTAAGTTGGTACAGAA 3720
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3801 CCTGTCACAGACCCCTGGAAGAGAAATTAGTATGATGACTTACTAAGTTGGTACAGAA 3860
QY 3721 CAGAAACCTTAAAGCAGTACGCAAGTCGGAAGAGCCATACGCGTTTCAAGTCTGAT 3780
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3861 CAGAAACCTTAAAGCAGTACGCAAGTCGGAAGAGCCATACGCGTTTCAAGTCTGAT 3920
QY 3781 GAACAGCATGGGCTGAGGAAAAAGGCTCAAAAGAAATATATTGAAATGAAGTGA 3840
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3921 GAACAGCATGGGCTGAGGAAAAAGGCTCAAAAGAAATATATTGAAATGAAGTGA 3980
QY 3841 CAGAAATAGTCCGCCAAAAAAGGGTAAAGAGGCCGACCAAAACCTTGTGGTGAAGT 3900
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3981 CAGAAATAGTCCGCCAAAAAAGGGTAAAGAGGCCGACCAAAACCTTGTGGTGAAGT 4040
QY 3901 ACACCAAAAGAAAGCCCAATGAAATCTTAAAAAAGAAACCAAAATCTTGA 3960
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4041 ACATGAAAGAAAGAGCCCAATGAAATCTTAAAAAAGAAACCAAAATCTTGA 4100
QY 3961 CCTCAGCAGCAGAGGAGAGAGAAAGAAAGCAAGTGAATACGGAACAGAG 4020
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4101 CCTCAGCAGCAGAGGAGAGAGAAAGAAAGCAAGTGAATACGGAACAGAG 4160
QY 4021 TCCAAAGCAACAGCAGCAGAGTGTCAAGAGAGCAGACAGAGAGAGAGATCTCTGAA 4080
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4161 TCCAAAGCAACAGCAGCAGAGTGTCAAGAGAGCAGACAGAGAGAGAGATCTCTGAA 4220
QY 4081 TCTAGTGCATTTGAATCCACAGAGTCCACACAGAGAAAGAGAGAGAGAGAGAG 4140
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4221 ACAAGTGCAGTTGAATCCACAGAGTCCACACAGAGAAAGAGAGAGAGAGAGAG 4280
QY 4141 ACGGCATCAACATCAACCAAAAAA 4169
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4281 GCACCATCAACATCAACCAAAAA 4309
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

RESULT 5
LOCUS AK122414 5944 bp mRNA linear ROD 15-MAR-2003
DEFINITION Mus musculus mRNA for mKIAA0979 protein.
ACCESSION AK122414
VERSION AK122414.1 GI:28972559
KEYWORDS FLI CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1

```

```

AUTHORS Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S.,
         Nakajima, D., Nagase, T., Ohara, O. and Koga, H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
JOURNAL gene: II. The complete nucleotide sequences of 400 mouse
REFERENCE KIAA-homologous cDNAs identified by screening of terminal sequences
          of cDNA clones randomly sampled from size-fractionated libraries
          DNA Res. 10, 35-48 (2003)
          (bases 1 to 5944)
AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
          Institute, Laboratory for Genome Informatics: 2-6-7
          Kazusa-kametari, Kisarazu, Chiba 292-0818, Japan
          (E-mail: mousekazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
          corporation; cDNA full insert sequencing; Kazusa DNA Research
          Institute; cDNA library construction, clone selection and 5'-
          3'-end one pass sequencing.
FEATURES
source
1..5944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbx00002"
/lisue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..5944
/gene="mKIAA0979"
<1..3577
/note="KDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=2
/evidence="not_experimental"
/product="mKIAA0979 protein"
/protein_id="BAC65696.1"
/db_xref="GI:28972560"
/translation="FDLILEYINSHLLSVLPQLEFKNSDNPEFLQYLLAKM
FGAKSELASOKRPMOCTLGFDNIHPIRECYRASHCLMNPDLAKDTLELYV
RSHDEEATRHQVIVSYIAKADLLVNDHLNLTRETTDKRVRKREAMEGLAET
YKYSLSAAGDAKQISWYDKLHLHYONSIDRLIVERIFQYVNPKNALTEER
MKCLYLVATLDLNAVKALNEMWKCQNLRHQVNDLIDIKQPKDAVSKALFSVM
ITRNLPDGRKADDPFKKFTFOVLEDEKIRKQLEALVSPSCSKQAEKVRTEKIGN
PKOPTNPELEMTKILIRAPVHIDTESISALIKOVNSIDETADDEGVTDOAIR
AGLEILKLSPTNPIRSHAFESBLIACIKMDDEKVAALQIETKNSKEEDPH
IRSALEPVLHRSKSGPPROKATVILHCHAIIFSSKETQPAQIEPLHSLDPSNLEH
ITPLVTIGHILAPDOFAAPLKLIVATFIVDLIMNRLPKRTTKLWNPDEVSPE
TPWKIOAIKMWYRWLIGMKNHNSKSTSLRLITLILSDGDLTQDGKISKDMSRLR
LAAGSAIYKLAQPCYHEITILEOYOLALAINDCYOYRQVFAOKLHGLSRLRLPL
EYMAICALCARKDPKERRAHAROCILVNTVRRREYLKQAAVSEKLSILPEYVPT
IHLAHDPDYKVVODI BOLKDYKECLMPEYLEILMAKNENSHAFIRKVENYIKQKDA
OGPDPTKMKNEKLYTCVDVAMNIMSKSTYSLESKPDVLPARFTQDPKNSNTKNY
LPEPKSFTPTPKPKTAVILGAVNPLSLSAGSOSQTSRMTVSNASSNPSSBGR
IKGRDLSSEMSSENYDMSPLPGKSKDEEDLSLEKPRSRKAPVDPDEKL
GMDLITLIVQEQPKSGORGRKRTASDSDEQMPDEERHKEELLENDQNSPK
GKRGKRPPLGGGTGSKKEPMTSKKKKKKTIYPPVVDDEDEEIOGTEHKSXSKO
HRTSKRAQGRASPESTSAVESTQSTPQGRGRSPKAPSPSPKIRGRSKQVATKE
NDSEEMOVLAISPSVSDTTQGEDEEDISVGNVRRRSKRERR"
BASE COUNT 1892 a 1151 c 1251 g 1650 t
ORIGIN
Query Match 67.5%; Score 2816.6; DB 10; Length 5944;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 3045; Conservative 0; Mismatches 354; Indels 6; Gaps 1;
QY 765 CTTTGACTTAATTTGAGCTTACAAATGTATGATGCTTCTGCTCTGTTTACC 824
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 CTTTGATTTGATCTTGAGACTCTACAACTTATACACCCATTGCTGCTTCTGCTGCC 60
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 825 CCAGCTGAATTTAAATTAAGAGCAATGATATGAGAGAGCCGCTACAAAGTTGTTAACT 884
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 CCAGCTTGAGTTTAATTAAGAGCAATGATATGAGAGAGCCGCTACAAAGTTGTTAACT 120
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

QY	885	ACTGGCAAAATGTTTGGGGCAAAAGATTCAAGATTTGGCTTCTCAAAACAAGCCACTTGG	944
Db	121	ACTGGCGAAAATGTTTGGGGCAAAAGATTCAAGATTTGGCTTCTCAAGAAAACCCCTCTGG	180
QY	945	GCAGGCGTACTTGGGCGAGTTTAAATGATATCCATGTACCAATCCGCCCTGGAAATGTGTAA	1004
Db	181	GCAGTGCCTACTGGGCGAGTTTAAATGATATCCATGTACCAATACGCCCTGGAAATGTGTAA	240
QY	1005	ATTTCCTAGCCATTGCTCATGAAACCATCTGATTTTAGCAAAAGACTTTACAGATATCT	1064
Db	241	ATTTCGCCAGCACTGCTCATGAAACCATCTGATTTTAGCAAAAGATTAAACAGATATCT	300
QY	1065	TAAAGTGGGTCACATGACGCTGAGGAAGCTATTAGCATATGTATTGTGTCAATAGT	1124
Db	301	TAAAGTGGGTCACATGATGCTTGAGGAAGCTATTAGCATATGTATTGTGTCAATAGT	360
QY	1125	TACAGCTCTAAAAAGATATTCTTCTGTGTAATGATATGATACCTACTATTATTTTGTGAGAGA	1184
Db	361	AACAGCTCTAAAAAGATATTCTTCTGTGTAATGATATGATACCTACTATTATTTTGTGAGAGA	420
QY	1185	GAGAACATTTAACAACAAGCATGGAGATATGCCAAAGAACCATGATGGGACTTGGCCCAAT	1244
Db	421	GCGAACATTTGACAACAACGCTGGAGAGTATCGTTAAAGAACCATGATGGGACTTGGCCCAAT	480
QY	1245	TTATTAAGAAATATGCTTTTACAGTACGACGCTGGAAAAATATCTGTCAAAAACAGATAGCATG	1304
Db	481	TTATTAAGAAATATCTTTTACAGTACGACGCTGGCAAAAGATCTGTCAAAAACAGATATCTGT	540
QY	1305	GATCAAGACAAATTTGCTACTATATATATTATCAAAATATGATTGATGATGACTACTTGT	1364
Db	541	GCTCAGAGCAAGAGCTGCTGCTATATATCTATACAGAACGATTCACAGCCGCTACTTGT	600
QY	1365	TGAACGGATCTTGGTCAATACATAGTGTCCACAAATTAAACTATAGAAACGAGATGAA	1424
Db	601	TGACGGATCTTGGTCACTACATAGTGTCCACAAATTGGAACGAGGAAACGAGATGAA	660
QY	1425	ATGCTTATATTACTTGTATGCGACACTGATTTAAATGCTGTGAAGCATTTGAATGAAAT	1484
Db	661	GTGCTTATATTACTTGTATGCGCTACACTGATGATGATGCTGTGAAGCATTTGAATGAAAT	720
QY	1485	GTGGAATATGTAAAAATTCGCTCCGACATCAAGTAAAGATTTTGGTTCACCTTGATTAGCA	1544
Db	721	GTGGAATATGTAAAAATTCACCTCCGACATCAAGTAAAGATTTTACTGTTTAAATTAACA	780
QY	1545	ACCCAAAACAGATCGCAGTGTCAAGAGCCATATTTTCAAAAGTGAATGTTATTACAGAAA	1604
Db	781	ACCAAAAACAGATCGCAGTGTCAAGGCCATATTTTCTTAAATGATGTTATCACAGAAA	840
QY	1605	TTTACCTGATCTCGGTAAAGSCTCAGGATTTCATGAAGAAATTCACACAGTGTTAGAGA	1664
Db	841	TTTTCGCAATCTCGGTAAAGGCCCAAGACTTCATGAAGAAATTCACACAGTGTTAGAGA	900
QY	1665	TGATGAGAAATTAAGAAAGCAGTTAAGAAATGTTGTTAGTCCACATCTGCTCCGCAAGCA	1724
Db	901	TGATGAGAAATTAAGGAACAGCTGGAAGCCCTTGTTAGTCCCAATGTTCTCTCCAGCA	960
QY	1725	GGCTGAAGGTTGTGTGCGTGAATTAAGTAAAGTTGGCAACCCCAACAGCCCTACAAA	1784
Db	961	GGCTGAAGAGCTGTGTGCGTGAATTAAGTAAAGTTGGCAACCCCTACAGCTTACAAA	1020
QY	1785	TTCCTTCTCGAATTAAGTAAAGTTTCCTTGGAGAGATACCTCTGTACATATGACATC	1844
Db	1021	CCCTTTCTTGGAAATGATCAAGTTTCTTGGAGAGATACCTCTGTACATATGACATC	1080
QY	1845	CGAATCTATACGCTCTCTTATTAACAAGTAAACAATATCATATAGATGGAACAGCAGATGA	1904
Db	1081	TGAATTCATCATGCTCTTATTAACAAGCAAGTAAATTAATCATATATGGAACAGCAGATGA	1140
QY	1905	TGAAGATATGAGGCTGTCCACTGATCAAGCCATCAGACAGGCTTGAATCTTTAAAGT	1964
Db	1141	TGAGATATGAGGCTGTCCACTGATCAAGCTATATAGGCGACGCTTGAATCTTTAAAGT	1200

OY	1965	ACGTCATTATACAACTCCCACTCATTTGCATTTCGTCTGAACAATTTTGAATCATTAAGTGCG	2024
Dd	1201	GCTGGCTTTTAGCGCACCCCTATCTCCTTTTCACTCTGCTGAGAACGTTTGAGTCCCTTCTTGCGC	1260
OY	2025	TTCGTCGAAAATGCATGATGATAAAAAAGTATGCAAGAAGCTCAGCTACAAATTTTCCAAAACAC	2084
Dd	1261	TTGTCGTGAAGATGATGATGAGAAGGTGGCAAGAGCTGCCTACTACAAATTTTCCAAAACAC	1320
OY	2085	AGGAAGCAAATTTGAAGAGGATTTTTCCACACATCAGATCAGCCCTTGCTTCCGTWTACA	2144
Dd	1321	AGGAAGCAAAATTTGAAGAGGATTTCCCACATCAGATCAGCCCTTGCTTCCGTACTACA	1380
OY	2145	TCACAAATCTAAAAAAGGAGCCCCGCCGTCAGAGCCAAATATGCAATTCATCTEATATCCAGC	2204
Dd	1381	TCACAAATCTAAAAAAGGAGCCACCCTGGCAGAGCCAAATATGCAATTCATCTEATATCCAGC	1440
OY	2205	GATATTTTCTAGTAAAGAGACCAGATTGTGCACAGATATTTAGGCTCTGCATAAGAGCCT	2264
Dd	1441	CATATTTTCTAGTAAAGAACCCAGTTTGCACAGATATTTAGGCTCTGCATAAGAGCCT	1500
OY	2265	AGATCCAAAGCAACCTGGAAACATCTCATTAACACATTSGTATCTATTSGTATATTTGCTCT	2324
Dd	1501	AGATCCAAAGCAACCTGGAGACATCTTATTAACCCCTGCTGCTACTATTSGTATATTTGCTCT	1560
OY	2325	CCTTGCACTATCATTTTGTGCTGCTCTTGGAATCTTGGGTAGCTACTTTTCATTTGTGAA	2384
Dd	1561	TCGTGCACCTGATCAGTTCCGTGCTGCCCTGTAAGTCTTTGGTGGCACTTTCATTTGTGAA	1620
OY	2385	AGATCTTTCATGAATGATGCGCTTCCAGSGAAAAACAACTAATCTTGGTGTCCAGA	2444
Dd	1621	GGACCTCCTCATGAAATGACCGGCTTCCAGSAAAAACAACTAAGCTTTTGGTGTCCAGA	1680
OY	2445	TGAAGAACTATCTCCTTGAGACAATGCTPAAAAATTCAGCCTTTTAAAAAGATGGTTCGATG	2504
Dd	1681	TGAGAGATCTCACCCTGAGACAATGCTPAAAAATTCAGCCTTTTAAAAAGATGGTTCGATG	1740
OY	2505	GCTACTTGGAAATGAAAAATTAATCAACGTAATTCAGGAACCTTCTACCTTAAGATTGCTAAC	2564
Dd	1741	GCTACTTGGAAATGAAAGATTAATCAACGTAATTCAGGAACCTTCCACCCTCAATTAATCAAC	1800
OY	2565	AACAAATATGATATGATGATGAGACCTTGACAGAACAGGGGAAAAATTTGTAACCAAGATAT	2624
Dd	1801	AACGATCTGCAATGATGATGAGGAGATTITGACGAACAAAGGAAAAATTTAETAACCAAGATAT	1860
OY	2625	GTCACGCTGAGACCTTGCTGCTGGGAGTGCATTTGTGAAGCTGGCACAGAACCCTGTTA	2684
Dd	1861	GTCACGCTGAGACCTTGCTGCTGGGAGTGCATTTGTGAAGCTGGGACAGAGCCCTGTTA	1920
OY	2685	CCATGSAATCATCATATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATG	2744
Dd	1921	CCAGAGATCATTTACACAGGAGCAGTACAGCTGTGTGCATTATCCAACTAATGATGAGTG	1980
OY	2745	CTATCAAGTAAAGCAAGTGTGTGGCCAGAAACCTTACAAAGGCGCTTCCGCTTAACGGCT	2804
Dd	1981	TTATCAAGTCAAGGAGGCTGTCCGTCAGAAACCTTCACAAAAGCGCTTCCGCTTAACGGCT	2040
OY	2805	TCACATTGAGTATATGGCAATCTGTGCCCTTTGTGC AAAAGATCCTGTAAAGAGAGAG	2864
Dd	2041	TCCTCTTAGTACATATGGCCATCTGTGCTCTTTGTGCC AAAMACCTGTGAAGAAGAGCG	2100
OY	2865	AGCTCATGCTTGGCAATGTTTGGTGAAAAATATTAAGTAAAGGGGGGAGTATCTGAAGCA	2924
Dd	2101	AGCCCATGCTAGACAGTCTTGGTGGAAGACATCAGTCTGTAGGAGGGAGTACCTGMAAGCA	2160
OY	2925	GCATGCAAGCTTATAGTAAAAATTAATGTCTCTTAC CAGAGTATGTTGTTCATATAC	2984
Dd	2161	GCATGCAAGCTTATAGTAAAAATTAATGTCTCTTCTTCA CAGAGTATGTTGTTCATATAC	2220
OY	2985	AATTCACCTTTTGGCACAATGACCCAGATATATGCTAAAGTACAGATATTTGAACAACTTAA	3044
Dd	2221	AATTCACCTTTTGGCACHVTGACCCAGATATATGCTAAAGTATAGATATTTGAACAACTTAA	2280
OY	3045	AGAGTTTAAAGAAATGCTTTTGTTGTTTCTGGAATATTTAATGCTTAAAAATGAATAA	3104

Db 2281 AGATGGAAGATGGCTTGTGTTGTCGAGATATGATGGCTAAATATGAAAAACA 2340
QY 3105 CAGTCACGCTTTATCAGAAAGATGTAGAAAAATATTAACAACAAACAGATCCCAAG 3164
Db 2341 CAGCCATGATTTATCAGAAAAATGTTAGAAAAATATTAACAACAAACAGATCCCAAG 2400
QY 3165 ACCAGATGATGCAAAAAATGAAAAATGCTACTGTGTGATGTTGCCATGAAATAT 3224
Db 2401 ACCAGATGATGCAAAAAATGAAAAATGCTACTGTGTGATGTTGCCATGAAATAT 2460
QY 3225 CATCATGTCAAAAGACTACTACATAGCTTTGGAATCTCCTAAAGACCCGGTACTACAG 3284
Db 2461 CATCATGTCAAAAGACACACATAGCTGAGGATGCTCCTAAAGACCCGGTACTACAG 2520
QY 3285 TCGTTTCTTCACTACCACTGACAGAAATTTCACTAACACCAAAATTTCTGCTCTCGA 3344
Db 2521 TCGTTTCTTCACTACCACTGACAGAAATTTCACTAACACCAAAATTTCTGCTCTCGA 2580
QY 3345 AATGAATCATTTTCTACCTCTGGAAGAAACCTAAACACCAATGTTCTAGACCTGTA 3404
Db 2581 AATGAATCATTTTCTACCTCTGGAAGAAACCTAAACACCAATGTTCTGAGCAGTTAA 2640
QY 3405 CAAGCCTTTCTACAGACGAGCAAGCAATCTCAGACCAATCATCAGAAATGGAATCTG 3464
Db 2641 TTAGCCACTTTCATCAGACGAGCAAGCAATCTCAGACCAATCATCAGAAATGGAATCTG 2700
QY 3465 AAGCAATGCACAGCAGCACTCAATCCAAAGCTCTCTCGAAGAAATTAAGGAGCTTGA 3524
Db 2701 GAGCAATGCACAGCAGCACTCAATCCAAAGCTCTCTCGAAGAAATTAAGGAGCTTGA 2760
QY 3525 TAGTCTGAAATGATCAGAGTAAATGAAATGAAATGAAATGAAATGAAATGAAATG 3584
Db 2761 TAGTCTGAAATGATCAGAGTAAATGAAATGAAATGAAATGAAATGAAATGAAATG 2820
QY 3585 GAAAAAAGTGAACAGAGCAGCTCTGTAAGGCTGTAATTTGGAGAGCTAG 3644
Db 2821 AAAAAAAGTGAACAGAGCAGCTCTGTAAGGCTGTAATTTGGAGAGCTAG 2874
QY 3645 AGGCAAGAAAAAAGCCCGCTCAGACAGACAGAGGAAATTAAGTATGATGATCTGAC 3704
Db 2875 AAGTCGAAAAAAGCCCTGACAGACCTGAGAAATTAAGTATGATGATGATGATGATG 2934
QY 3705 TAAGTGTGACAGAGCAAGCAAGCTTAAGGCACTGACGAGTGCAGAAAAAGAGCCATAC 3764
Db 2935 TAAGTGTGACAGAGCAAGCAAGCTTAAGGCACTGACGAGTGCAGAAAAAGAGCCATAC 2994
QY 3765 GGGTTCAGATCTGATGAACAGAGTGGCTTGAAGAAAGAGGCTCAAGAAAGATATAT 3824
Db 2995 AGCTTCAGACTCAGAGAGAGAGAGTGGCTTGAAGAAAGAGGCTCAAGAAAGATATAT 3054
QY 3825 AGAAAAATGAAGTGAACAGATATGTCGCCAAAAAAGGTTAAAGAGCCGACACCAAA 3884
Db 3055 GGAATATGAGATGAGACAGAACAGCCACCAAAAAAGGCGAAAAAGAGCCGACACCAAA 3114
QY 3885 ACCCTCTGTTGAGAGTACACCAAAAAAGAGCCCAAGATGAAATCTTCTAAAGAAAGAG 3944
Db 3115 ACCCTCTGTTGAGAGTACACCAAAAAAGAGCCCAAGATGAAATCTTCTAAAGAAAGAG 3174
QY 3945 CAAAAAATATCTGAGCTCCAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4004
Db 3175 CAAAGAAAAAATCTTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3234
QY 4005 AATATCGAAGCAAGAGTCCAAAAAGCAAGCAGAGTGTCAAGAGAGACACAGAGAG 4064
Db 3235 AATATCGAAGCAAGAGTCCAAAAAGCAAGCAGAGTGTCAAGAGAGACACAGAGAG 3294
QY 4065 AGAGAAATCTCTGATCTAGTCAATTAATCCACACAGCTCCACACCAAGAAAGAGAG 4124
Db 3295 AGAGAAATCTCTGATCTAGTCAATTAATCCACACAGCTCCACACCAAGAAAGAGAG 3354
QY 4125 AGGAAAGCCATCAAAAAAGCCATCAGCATCACAACCAAAAAA 4169
|||||

Db 3355 AGGAAAGCCATCAAAAAAGCAAGCAGATCAGCAACACCCCAAAAA 3399
RESULT 6
AF294791
LOCUS
DEFINITION
Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION
AF294791
VERSION
AF294791.1
KEYWORDS
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 6744)
Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
Identification and characterization of SCC-112, a novel cell cycle
regulated gene in human cancer cells
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 6744)
Kumar,D. and Kasid,U.
Direct Submission
Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
Washington, DC 20007, USA
FEATURES
source
1..6744
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="4"
1..6744
/gene="SCC-112"
232..4125
/gene="SCC-112"
/codon_start=1
/product="SCC-112"
/protein_id="AA082347.1"
/db_xref="GI:21951802"
/translation="MKRIKAKVYKTTMDQSEDEKQOYLPLALHASEFFLRNPK
DVRLLVACCLADIFRIYAPAPYTSDDKLDKIDFLITRDLKGLIEDKSPQNFYFL
ENLAWKSYNICEFELEDNEHIFQLFRTPESVYNNHNKQVMHMDLWSSILMEDG
VTOELGSLILNIPAKHKNINOSFDLAVYLKRTYQTEACIANFNQVLYVGRSSV
SDLSHFVPLIOELFIDPHLLSVAPDLFRKLSDEGRLEAVVLLAKLGRSRA
LTONRPLMOCPGRENDFHVPRLRESYVFASICLNHNDLAKLLEYIKVSHDEE
ATRHDIYVITTRAKDLALVNDQLLEFVREBRLDKRMKVRKAMGLAQIKTCLH
GAGKAAERKSWIKDKLHIYQNSIDKLVLEKIFAYLVPHNLETERMKCLYL
VASLDPNVAKALNEMKCONMLSHVRELLDHPQTSBANSAMFGKLMATKANLPD
PGKAOFVKKFNQVLDGDEKLSOLELLSPSCSKQADICVLEIRKLIANPQPNP
PLEMKVFLERLAPVHIDSEALISAVKLANKSTEGRADEEESVDPVIRSGLELK
VLSFHPITPSHASEYTESLQCLRMEDDKVAEALDIPFNTGKITFDLPQIRSTLIP
LHOKAKRGTPIHOKAIVCHITFNKEVQLAQIFELSRSLNADVPOLITPLVSL
GHSLMAPDQFASPMKSVANFIVKDLNDRSTGRKNGKLSMPDEVSPEVLAQVA
IKLVMLGLMKNNOSKANSSTLRLLSAMLVSGDLEGRKLSKSMSPRLTLAGSAI
MKLAOEPCHYEITTPPOFOLALVIDECYQVYRFAOKLHALVLLPLLEMAIFA
LCAKDPVREBRARHAROCCLKNLSIREYIKQPMATEKLSLIPVYVYVYMLHIALD
PORTRSQVDQDLDRICECLMPELVMTNENNSAFMKMAENIKLITRDAOSSPDEK
TNEKLYTCVDVALCVINRSALCNADSPDVPVPMKFFQPEKDFCNDSTYISETRV
LILTGKRPAGVIGAVNKPISATGRKPYVRSGETEGSINIVNSLNPSTGRNSRS
SEAAETGVSENEENPVRIISVTPVKNIIDVKNKEISDQATOGNISSDGKRTVYAA
GSPGQKTEDEKVDGSPKSPRGRPKSPSOGNATKNDLNPINKGRRAAVG
OESPGLELGNMKAPKLPDLPARKKAPAEQIDLPK"
BASE COUNT 2156 a 1243 c 1400 g 1945 t
ORIGIN
Query Match 43.5%; Score 1815.6; DB 9; Length 6744;
Best local Similarity 69.7%; Pred. No. 0;
Matches 2457; Conservative 0; Mismatches 1069; Indels 0; Gaps 0;
QY 32 GAAAAATTCATATCCGCTGGGCTCAAGAAATATCAGATTAATATATTAAGAGGACA 91
Db 173 GGAAGATCGCTTACCTCCGGGGGTAAAAAGAGATCCACCAAGAACACACAGAGAGACA 232
QY 92 TGGTAGAGGATTAAGGATGTTGTGAAAACTTTATGATATGAGCAGAGACTGTGAG 151

Db 233 TGATCAAAAGCCGTGAAGATGGTAGGAAACCTTTATGATATGATGATGAGCAGCTCAGAG 292
Qy 152 AAGAAAGAGGCTTATTTAAACCGTGTACACTGCTGCAGATTTTTTCCAGC 211
Db 293 ATGAAAAACACAGATATCTCCACAGCTTGACATCTGCATCTGAATCTTCCACAGA 352
Qy 212 ATCTGTGAAGANGTTCCTTACTGAGTGCCTGCTGCATATTTTTCAGATTT 271
Db 353 ACCCAATTAAGATGTGGTCTCTCTGTAGCATGTGTTGGCTGATATCTTGGTACT 412
Qy 272 ATGCTCTGAAGCTCTTACACATCCCTGATTAACCTAAGAGATATATTTATGTTTAA 331
Db 413 ATGCCCCAGAGCTCCATATCTCTCCATGAATAACTTAAGACATATTTTGTATTA 472
Qy 332 CAAGACAGTTGAGGGGCTAGAGGATTAACAAAGGCCAATTCATAGTATTTTAT 391
Db 473 CCAGACATTAAGAGTTGGAGAGATACAAAGAGTCCACAGTTAATGATATCTTTTAT 532
Qy 392 TACTGAGACATTCCTGGGTCAAGTCAATTAACATATGCTTTGAGTTAGAAGATGCA 451
Db 533 TATTAGAAATTTAGCTGGGTAAATCATATTAACATCTGCTTGAATTTGAAGATGCA 592
Qy 452 ATGAAATTTTCAACCAGTATACAGAACCTTATTTTCAGTTATTAACATGGCCCATC 511
Db 593 ATGAAATTTTATTCAGCTTTTGAACCTCTCTCTCAGTATCAACAAATAGCCCAATA 652
Qy 512 AGAAAGTCCATATGACATGATGACCTTATGAGCTTATTTGTAAGGTGATACAG 571
Db 653 AGAAGTCAAAATGACATGCTAGTTGATGAGTTCATCATCATGAGAGTGATGAG 712
Qy 572 TGTCTCAGAGCTTTTGGATACGGTTTAAATCTGTGTAACCTGCTCATAGAAATTTAA 631
Db 713 TTACCAAGAAATTTATGGGCTCCATCTTATTAACCTCATCTCCACATAGAACTTAA 772
Qy 632 ACAAGCAAGAAATTTGGCAAGGCTTACTACAAAGAGAGGACCTCAAGTATTTAGC 691
Db 773 ATAAACAGTCCCTTGACCTTGCAAAAAGTCAATTTGAAAAAGAACACATCCAGCTATTGGG 832
Qy 692 CATATATTAACCACTTTTAAATCAGTTTGCATGCTGGGAAACATCTATCAGCAGAT 751
Db 833 CATGATTTGCCAATTTTTCATCAACAGTCTGCTGCTGGAAAGATCATCAGATGATTT 892
Qy 752 TGTCAAGCATGCTTTTACTTAATTTTGGAGCTCTACAAATTTGATGTCATTTGCTGC 811
Db 893 TGTCAAGCATGATTTATCTGATCTGAGGAACCTTTTGTCAATATATCTCTGTATTTAT 952
Qy 812 TCTCTGTTTACCCAGCTTGAATTAATTAAGAGCAATGATGATGAGGAGCCCTAC 871
Db 953 TATCCGTCATGCGACAGCTTGAATTTCAAACTTAAGAGCAATGATGAGAGGAGCATTTAG 1012
Qy 872 AAGTTGTTAACTACTGGCAAAAATGTTGGGGCAAGAGATTCAGAAATTTGGCTTCAAA 931
Db 1013 CTGTTGTTGACCTTACTAAATGTTTGGCTCCAAAGATTTCTGATTTGGCAACAGAA 1072
Qy 932 ACAAGCCACTTTGGAGAGTACTTGGCAGGTTAATGATATCCATGTACCAATCCGCC 991
Db 1073 ATCGCTCTTTGGCAATGTTTCTTGAGCAGATTTAATGATATCTATCTGTGAAT 1132
Qy 992 TGAATGTGTGAATTTGCTAGCCATGTGCTCATTAACCATCCCATTTAGCAAAAGCT 1051
Db 1133 TAGAAGTGTGAATTTGCCAGTCAATGTTTAAATACCCACATTTAGGGAAGATC 1192
Qy 1052 TAACAGATATCTTAAAGTGAAGTCAATGACCTTGAGGAAGCTTTTGAAGTATGTTTA 1111
Db 1193 TCACGATATTTTAAAGTTAGATCAATGATCAGAGAAAGACTTTTGTCTATGATGCA 1252
Qy 1112 TTGTCTCAATAGTACAGCTGCTAAAGAGATATTTCTTGTGCTCAATGATCACTTACTTA 1171
Db 1253 TTGTACTATATAATACAGCTGCAAGAGGAGCTGGCTTGTATTAATGCTAGCTGTG 1312
Qy 1172 ATTTGTGAGAGAGAACTTTAGCAAAAGCATGAGAGTACGCAAAAGAGCATGATGG 1231

Db 1313 GCTTTGTAAAGGAAAGAACTGATTAACCGGTGGCGAGTAAAGAAAGAACTATGATG 1372
Qy 1232 GACTTGCCCAATTTTATAGAAATATGCTTTACAGTCCAGCTGGAAGAAAGATCTGCA 1291
Db 1373 GTCTGGCTACCTTTATAGAAATATCTCTTCAATGGTGAAGCAAGAAAGAGCTGAG 1432
Qy 1292 AACAGATGACATGATCAAGCAAAATGCTACATATATATTAATCAAAATAGTATGATG 1351
Db 1433 AGAAAGTCAAGCTGATAAGAGCAAACTTTGCAATTTATTAATCAAGAAAGCATTTGACG 1492
Qy 1352 ATCGACTCTTGTGAAGCGATCTTTGCTCAATACATGATGTTCTTCACAAATTTAACA 1411
Db 1493 ACAACCTGTGTAGAGAAATCTTGTGCTAGTATCTTCCACCAACCTGGAAGAG 1552
Qy 1412 CAGAAAGGATGAATATGCTTATTAATTAATGTCACACTGATTTTAAATCTGTAAG 1471
Db 1553 AAGAGAGATGAATATGCTTATATTAATTAATGCTAGTTTGATGCAAAATCTGTAAG 1612
Qy 1472 CATTGATGAATGTGAATGTCAAAATCTGCTCCAGCATCAAGTAAAGATTTGCTTG 1531
Db 1613 CTCTCAACGAATGTGGAAGTGTCAAGAACATGCTCGAGGACATGACGCAACTATGG 1672
Qy 1532 ACTTGATTAAGCAACCCAAACAGATGCGCAGTCAAGGCGCATATTTTCAAAAGTATG 1591
Db 1673 ATTTGCACAGACGCTTACATCAGAGCTAATCTGCTTCCATGTTTGAAGAACTGATGA 1732
Qy 1592 TTATTAACAAGAAATTTACCTGATCCGTGAAGGCTCAGAGATTTGATGAAGAAATTCAC 1651
Db 1733 CCATAGCAAGAAATTTGCTGACCCGGGAAGCAAGATTTTGTGAAGAAATTTAAC 1792
Qy 1652 AGGTGTTGAAGATGATAGAAATTAAGAAAGCACTTGAAGTACTTTTGTGCTCAACAT 1711
Db 1793 AGGTCTCGGGAGTATAGAAACTCGCTCAGTGTGAATTTATTAATTTAGCCAACT 1852
Qy 1712 GCTCTGCAAGAGGCTGAAGTGTGTGCGTGAATTAAGAAAGTGGGCAACCCCA 1771
Db 1853 GTTCTTGCAACAGACAGATTTTGTGTGAGAGAAATGCCCCGGAACCTTCCAAATCTTA 1912
Qy 1772 AACAGCTTACAAATCTTCTCGAAGATGATCAAGTTCCTTGTGAGAGATGACCTG 1831
Db 1913 AGCAACCAAAATCTTCTTGAAGATGATCAATTTCTGTGGAAGAAATTCGACCTG 1972
Qy 1832 TGCATATGATACGGAATCTATCAGTGTCTTATTAACAAGTGAACAAATCAATAGTG 1891
Db 1973 TGCATATGATTCAGAAACCAATAGTGACATGAGAAATTAAGTAAAGCAATAGAGG 2032
Qy 1892 GAACAGCATGATGAAGATGAGGTGTCCAACTGATCAAGCATCAGACAGAGCTGTG 1951
Db 2033 GGACAGCATGATGAAGAGAGGTGTAACTCCAGATACAGCTATCCGTTCCAGSACTTG 2092
Qy 1952 AACTGCTTAAGGTACTCATTTACATCCATCTCATTTTCAATTTGCTGAACATTTG 2011
Db 2093 AACTCTTAAGGTCTGCTTTACACATCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2152
Qy 2012 AATCATTAAGCTTGTGTGAAGATGATGATGAAGAAAGTGAAGCAAGCTGCTGCTGCTG 2071
Db 2153 AGTCTCTTGTACAGTGTGCTGAAGATGAGGATGAGAAAGTGAAGCAAGCTGCTGCTGCTG 2131
Qy 2072 TTTTCAAAACAGAGAAATTTGAAGAGATTTTCCACACATCTGATCAGCTGCTGCTG 2131
Db 2213 TTTTGAAGAAATGAGGTACAAATAGAAACAGACTTCCACAGATACGATGACCTTAA 2272
Qy 2132 TTCTGTTTAACTCAAAATCTTAAGAAAGACCCCGCTCAAGCCAAATATGCAATTC 2191
Db 2273 TTCCATTTTACATCAAAAGAAAGAGGGGTACTCCACACAGCAAAACAGGCTGTC 2332
Qy 2192 ATGTATATCAGCATATTTTCTAGTAAAGAGACCAGTTTGCACAGATATTTGAGCTC 2251
Db 2333 ACTGTATACAGCCCTTATTAACAATTAAGAGTCCACACTGCAAGATTTTGTGAGCCAC 2392
Qy 2252 TGCATTAAGAGCTTACATCCAAAGCACTGGAACATCTCATTAACCATTTGTTACTATG 2311
Db 2393 TCAGTAGAGTCTGAATCTGATGTGCCAGAAACAATTAATCAATTCATTAATTTGCTATGG 2452

OY	2312	GTCAATATGCTCTCCCTTCACCTGCATCAATTTGCTGCTTGGAAATCTTGGTAGCTA	2311
Db	2453	GCACATTTTCATGCTTTCACACAGATCGTTGGTCTCCCAATGAAATCTGTAGACCA	2512
OY	2372	CTTTCATGTGAAAGATCTTCATGATGATGCGGCTTCCAGGGAAAAAGCAACTAAC	2431
Db	2513	ATTTTATTTGTAAAGATCTGCTAATGATGACAGGTCAACAGGTGAAAGATGGAANAC	2572
OY	2432	TTTGGGTTCCAGATGAAAGAACTATCTCCTGAGACAAATGCTMAAATTCAGGCTATTAAA	2491
Db	2573	TGTGTTCTCCAGATGAAAGAGGTTTCCCTCGAAGTACTAGCAAGGTACAGGCATTTAAC	2632
OY	2492	TGATGCTTCAGATGGCTACTTGGAAATGAAAAATATCACAGTAAATCAGAACTTCTACT	2551
Db	2633	TTTCGTGAAGGTGGCTGTGGGTATGAAAAAACCAACACTCTAAATCTCCATTCACACC	2692
OY	2552	TAAATTTGCTTAACAACAAATATTGCAATAGTAGAGAGCTTACAGAACAGGGGAAAAATTA	2611
Db	2693	TTCCGTTTATTACAGCGAATGTTGGTTAGTAGAGGGGAGCTTACAGAGCAAAAGAGATCA	2752
OY	2612	GTTAAACCAAGATATGTCACGTCGTGAGACTTGTCTGCGGAGTCTAATTTGTGAAGCTGCAC	2671
Db	2753	GTAATATCGAATATGCTCGCTTGGGATAGCTGCTGGTAGTGCCATATGAAGCTTGCTC	2812
OY	2672	AAGAACCCTGTACCATGAATTCATTCATTAAGAAACATATACGATATGCTCATTTAGCTA	2731
Db	2813	AGGAACCTGTTTACCATGAATTTATTATCCCAAGAACATTTCCAGCTCTGTGCACCTGTGTA	2872
OY	2732	TCAACGATGAATGCTATCAAGTAAGACAAAGTGTTCGCCAGAAACTTCCACAAAGCCCTTT	2791
Db	2873	TTAATGATGATAGTGTACCAAGTAAGGCGAGATATTGTCTCAGAGCTGCATTAAGGACCTTG	2932
OY	2792	CCCGTTTACGGCTTCCACTTBAAGTATATGGCAATCTGCGCCCTTTGTGCAAAAGATCCTG	2851
Db	2933	TGAAGTACTCTCCCATTTGAGATATAGGAGATCTTTCCTTGTGTGCCAAAGATCTCTG	2992
OY	2852	TAAAGGAGAGAAGAGCTATCTAGGCAATGTTTGGTGAATAATATAATTAAGTAAAGCCGG	2911
Db	2993	TGAAGGAGAGAAGACACAGCCAGCAGACAACTTTACTGAAAAATATCATGTTATCCAGAGG	3052
OY	2912	AGTATCTGACAGACATGCACTGTTAGTGAATAATATTGTCTCTCTTACCAAGATATG	2971
Db	3053	AATACATTAACCAAAATCTATGGCTACTGGAATAATTTATCTACTGTTGGCTGAATATG	3112
OY	2972	TTGTTCCATTAACAACTTCCCTTTTGGCACATGACCCGATTAATGTCAAACTACAGGATA	3031
Db	3113	TAGTTCCATCATGATGATTCACCTGCTAGCCCATGATCCAGATTTTACAGATACCAAGTATG	3172
OY	3032	TTGAACAACCTTAAAGATGTTAAAGATGCTCTTGGTTGTCTGGAATATTAAATGGCTA	3091
Db	3173	TTGATACGCTTCGTGATATCAAAAGAGTCCATGTGTTCTGTAAGTATTAAATGACAA	3232
OY	3092	AAATGGAATAAAGTCAGCGCTTTTATCAGAAAGATGATGAAATATTAAACAACAA	3151
Db	3233	AGATGGAATAAATAGCCATGCTTTTATGAAGAAAGATGGCAGAGAACATCAAGTTAAACCA	3292
OY	3152	AAGATGCCCAAGGACACAGATGATCCAAATAATGAATGAAAACTGTACACTGTGTGATG	3211
Db	3293	GAGATGCCCAAGTCGCCAGATGAATCCAAAGAACAAATGAAAACTGTATACATATGATATG	3352
OY	3212	TTGCGATGAATATCATCATGTGCAAAAGAGTACTATACATACAGTTTGGAAATCTCTAAAGAC	3271
Db	3353	TGCGTCTCTGTTATTAATATGTAAGTATGCTTTGTGTGCATACAGATTCACCAAAAGCAC	3412
OY	3272	CGGATACACAGCTCGTGTCTTTCACCTCAACCTGCAGAAATTTAGTAAACCAACAAATTT	3331
Db	3413	CAGTCCTCCCAATGAATTTTATACACAACTGAAAAAGACTTCTGTATACATATGAAGTT	3472
OY	3332	ATCTGCTCTCGAATGAATCATTTTTCACCTCTCGAANAACCTTAACAACAACATGTTCT	3391
Db	3473	ATATTTTCAAGAGACACAGATCTTCTGTATACAGGAACCAAGGCTCTGTGAGATAC	3532

QY	3352	TAGGACGCTGTTTAACAAGCCACTTTCATCAGACGACCAATCTCAGACCAATCATCAC	3451
Db	3533	TAGGCGCAGTAAATTAAGCCTTTATCAGCAACGGGAGGAAACCCATGTTGAAGACACTG	35922
QY	3452	GAATGGAACGCTAAAGCAATGCAAGACACACCTCAAAATCCAAAGCTCTCCTGGAAAGTA	35111
Db	3593	GCATGACGACGGAAGCAATATTTAATGTAAATTCAGACCTGACACCTTCAACCGGAATTC	36522
QY	3512	AGGGGAGCGCTTGATAGTCTGAAATGATCAGATCAGTGAATGAAGA	3557
Db	3653	GATCAAGGGAACAGAGTTCAAGAGCAGCAGCAAACTGAGTTAGTGA	3698
RESULT 7			
HSU50533			
LOCUS	HSU50533	1852 bp	mrna
DEFINITION	Human BRCA2 region, mRNA sequence CG008.	Linear	PRI 27-NOV-1996
ACCESSION	U50533		
VERSION	U50533.1	GI:1531605	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1852) Couch,F.J., Rommens,J.M., Neuhausen,S.L., Bellanger,C., Dumont,M., Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L., Farid,L., Frye,C., Hattier,T., Janekci,T., Jiang,P., Kehrer,R., Leblanc,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y., Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M., Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T., Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H., Labrie,F., Stolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L., Tavtigian,S.V. and Simard,J.		
TITLE	Generation of an integrated transcription map of the BRCA2 region on chromosome 19q12-q13		
JOURNAL	Genomics 36 (1), 86-99 (1996)		
MEDLINE	96411650		
PUBMED	8812419		
REFERENCE	2 (bases 1 to 1852)		
AUTHORS	Simard,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular Endocrinology, CHUL Research Center, 2705, Boulevard Laurier, Quebec City, Quebec G1V 4G2, Canada		
FEATURES			
SOURCE	Location/Qualifiers 1..1852 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="13" /map="13q12-q13" /note="CG008; DSG numbers: D13S171 and D13S310" 609 a 348 c 379 g 505 t 11 others		
BASE COUNT			
ORIGIN			
Query Match	42.0%: Pred. 1754.4; DB 9; Length 1852;		
Best Local Similarity	98.0%: Score. No. 0;		
Matches 1817; Conservative	10; Mismatches 21; Indels 6; Gaps 5;		
QY	1064	TTAAAGTGTGATCAGATGACCCCTGAGAAAGCTATTAGACATGATTTATGTGCAATAG	11233
Db	1	TTTAAAGTGTGATCAGATGACCCCTGAGAAAGCTATTAGACATGATTTATGTGCAATAG	60
QY	1124	TTACAGCTGCTAAAAAAGATATTTCTTGTGTCATGATCACTTACTTAAATTTTGTGAGAG	11833
Db	61	TTACAGCTGCTAAAAAAGATATTTCTTGTGTCATGATCACTTACTTAAATTTTGTGAGAG	120
QY	1184	AGAGAACCTTTGACAAACGATGGAGAGATGACCAAAAGAACCATGATGGAGATTCGCCAAA	1243
Db	121	AGAGAACCTTTGACAAACGATGGAGAGATGACCAAAAGAACCATGATGGAGATTCGCCAAA	180
QY	1244	TTTATTAAGAAATATGCTTTATCAGTCAGACGCTGGAAGAAAGATGCTGCAAAACGATATCAT	13033

```

|||||
Db 181 TTTATAGAAATATGCTTTACAGTCACAGCTGGAAAAGATGCGCAAAACAGATACAT 240
QY 1304 GGATCAAGACAAATGCTACATATATATCAAAATAGATATGATGAGCTACTTG 1363
Db 241 GGATCAAGACAAATGCTACATATATATCAAAATAGATATGATGAGCTACTTG 300
QY 1364 TTGAGGAGATCTTGTCTCAATAGATGCTCTCACATTTAGAAATACAGAGATGA 1423
Db 301 TTGAGGAGATCTTGTCTCAATAGATGCTCTCACATTTAGAAATACAGAGATGA 360
QY 1424 AATGCTATATTTACTTGTATGCACTGATTTAAATGCTGTGAAGCATTTGATGAAA 1483
Db 361 AATGCTATATTTACTTGTATGCACTGATTTAAATGCTGTGAAGCATTTGATGAAA 420
QY 1484 TGTGAATGTCAAAATCTGCTCCGACATCAAGTAAGATTTGCTGTGATTTAGC 1543
Db 421 TGTGAATGTCAAAATCTGCTCCGACATCAAGTAAGATTTGCTGTGATTTAGC 480
QY 1544 AACCCAAACAGATGCCAGTGTCAAGGCCATTTTCAAAAAGTGAATTTATTAACAGAA 1603
Db 481 AACCCAAACAGATGCCAGTGTCAAGGCCATTTTCAAAAAGTGAATTTATTAACAGAA 540
QY 1604 ATTACCTGATCCGTTAGGCTCAGATTTCAATGAAATTCACACAGTGTAGAG 1663
Db 541 ATTACCTGATCCGTTAGGCTCAGATTTCAATGAAATTCACACAGTGTAGAG 600
QY 1664 ATGATGAGAAATTAAGAAACAGTTAGAACTTGTAGTCCAAATGCTCCGCAAGC 1723
Db 601 ATGATGAGAAATTAAGAAACAGTTAGAACTTGTAGTCCAAATGCTCCGCAAGC 660
QY 1724 AGGCTGAAGTGTGTGCTGTAATACTAAGAAATTTGGGCAACCCCAACAGCCTCAAA 1783
Db 661 AGGCTGAAGTGTGTGCTGTAATACTAAGAAATTTGGGCAACCCCAACAGCCTCAAA 720
QY 1784 ATGCTTCCGTAAGATGATGATTTCTCTTGGAGAGATAGCACTGTGCACATTAATA 1843
Db 721 ATGCTTCCGTAAGATGATGATTTCTCTTGGAGAGATAGCACTGTGCACATTAATA 780
QY 1844 CCGAATCTACAGTCTCTTATTAAACAATGAACAAATATGATGAGAAACAGCAGATG 1903
Db 781 CCGAATCTACAGTCTCTTATTAAACAATGAACAAATATGATGAGAAACAGCAGATG 840
QY 1904 ATGAAGATGAGGCTGTCACATGATCAAGCCATCAGAGAGTCTTGAACCTGTTAAG 1963
Db 841 ATGAAGATGAGGCTGTCACATGATCAAGCCATCAGAGAGTCTTGAACCTGTTAAG 900
QY 1964 TACGTCATTTAACAATCCATCTCATTTCTGCTGAACATTTGATGATTAATCTAG 2023
Db 901 TACGTCATTTAACAATCCATCTCATTTCTGCTGAACATTTGATGATTAATCTAG 960
QY 2024 CTTGCTGAATATGATGATGAAATAGCAGAGTGTGATCAATTTTCAAAAACA 2083
Db 961 CTTGCTGAATATGATGATGAAATAGCAGAGTGTGATCAATTTTCAAAAACA 1020
QY 2084 CAGAGACAAATTTGAAGAGATTTTCCACACATCAGATCAGCTTCTCTGCTTTTAC 2143
Db 1021 CAGAGACAAATTTGAAGAGATTTTCCACACATCAGATCAGCTTCTCTGCTTTTAC 1080
QY 2144 ATCAACAATCTAAAAAAGACCCCGG-TCAGCCAAATATGCTATTTGATATCAT 2202
Db 1081 ATCAACAATCTAAAAAAGACCCCGGCTCAAGCCAAATATGCTATTTGATATCAT 1140
QY 2203 GCGATATTTCTAGTAAGAGACCCGTTTGAACAGATTTTGAAGCTGTGCAATAAGC 2262
Db 1141 GCGATATTTCTAGTAAGAGACCCGTTTGAACAGATTTTGAAGCTGTGCAATAAGC 1200
QY 2263 CTGATCTCAAGCAACCTGAGACATCTCATACACATGCTGTTACTATTTGCTATTTGCT 2322
Db 1201 CTGATCTCAAGCAACCTGAGACATCTCATACACATGCTGTTACTATTTGCTATTTGCT 1260
QY 2323 CTCCTTGACCTGATCAATTTGCTCTCTTGAATCTTGGGTAGCTACTTTCAATTTGTG 2382
|||||

```

```

Db 1261 CTCCTTGACCTGATCAATTTGCTGCTCTTTGAAATCTTTGTAGCTACTTTCATTTG 1320
QY 2383 AAGATCTCTCAGTAATGATTCGCTTCCAGGGAAGAAAGCACTAATCTTGGTTCCA 2442
Db 1321 AAGATCTCTCAGTAATGATTCGCTTCCAGGGAAGAAAGCACTAATCTTGGTTCCA 1380
QY 2443 GATGAAGAAATATCTCTGAGACATGTCGCAAAATTCAGGCTTTAAATGATGTTTCA 2502
Db 1381 GATGAAGAAATATCTCTGAGACATGTCGCAAAATTCAGGCTTTAAATGATGTTTCA 1440
QY 2503 TGGCTACTTGGATGAAAAATATACACAGTAATCAGAACTTCTACTTAAG-ATTGCT 2561
Db 1441 TGGCTACTTGGATGAAAAATATACACAGTAATCAGAACTTCTACTTAAGSATTGCT 1500
QY 2562 AACACAAATTTTCATAGTATGATGAGACTTGAACAGAACAGGGGAAAAATTAAGTAACAGA 2621
Db 1501 AACACAAATTTTCATAGTATGATGAGACTTGAACAGAACAGGGGAAA--TTAGTAACAGA 1557
QY 2622 TATGTACGTCGTGAGCTTCTGCTGGAGTGTCTATTGTGAGCTGGACACAGACCTTG 2681
Db 1558 TATGTACGTCGTGAGCTTCTGCTGGAGTGTCTATTGTGAGCTGGACACAGACCTTG 1617
QY 2682 TTACCATGAATCATCATTTAGAACAAATTCAGCTATGTCGATTTGCTATCAACGATGA 2741
Db 1618 TTACCATGAATCATCATTTAGAACAAATTCAGCTATGTCGATTTGCTATCAACGATGA 1677
QY 2742 ATGCTATCAAGTAAGACAACTGTTT-6CCGAGAACTTCCAAAGGCTTCCCGTTTAC 2800
Db 1678 ATGCTATCAAGTAAGACAACTGTTTGGCCAGAAACTTCCAAAGGCTTCCCGTTTGC 1737
QY 2801 GGCCTTCACCTTGAATATGAGCAATCTGTCCCTTTGTGCAAAAAGATCTGTAAAGAGA 2860
Db 1738 GGCCTTCACCTTGAATATGAGCAATCTGTCCCTTTGTGCAAAAAGATCTGTAAAGAGA 1797
QY 2861 GAAGAGCTCATGCTAGGCAATTTGCTGAAAAATTAATGTAAGCGGAGCT 2914
Db 1798 GAAGAGCTCATGCTAGGCAATTTGCTGAAAAATTAATGTAAGCGGAGCT 1851

```

```

RESULT 8
BC039256
LOCUS 4252 bp mRNA linear PRI 06-NOV-2002
DEFINITION Homo sapiens, similar to androgen-induced prostate proliferative
shutoff associated protein, clone MGC:33757 IMAGE:5295109, mRNA,
complete cds.
ACCESSION BC039256.1 GI:24657778
VERSION 1
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4252)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genetics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT
NTH-MGC project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcapsb-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiroki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-sbpc.stanford.edu
Contact: (Dickson, Mark) mcdépxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNU at: <http://image.lnu.gov>
Series: IRAC Plate: 48 Row: k Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7705287.

FEATURES

source

Location/Qualifiers
1..4252
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:33757 IMAGE:5295109"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
71..1660
/codon_start=1
/product="Similar to androgen-induced prostate proliferative shutoff associated protein"
/protein_id="AAH39256.1"
/db_xref="gi:24657779"

CDS

BASE COUNT 1407 a 664 c 792 g 1389 t
ORIGIN
Query Match 35.1%; Score 1464.6; DB 9; Length 4252;
Best Local Similarity 95.5%; Pred. No. 2.3e-291;
Matches 1518; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

1 ATGGCTCATTTAAAGACTAGAGCCAAATGATGAAATTTACATATCCGCTGGGTCAAG 60
|||||
71 ATGGCTCATTTAAAGACTAGAGCCAAATGATGAAATTTACATATCCGCTGGGTCAAG 130
61 GAAATATCAGATTAATAATCTAAAGAGAGATGTGAGACGATTAAGATGGTGTGAAA 120
|||||
71 ATGGCTCATTTAAAGACTAGAGCCAAATGATGAAATTTACATATCCGCTGGGTCAAG 130
131 GAAATATCAGATTAATAATCTAAAGAGAGATGTGAGACGATTAAGATGGTGTGAAA 190
121 ACTTTATGATGATGAGACGAGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 180
191 ACTTTATGATGATGAGACGAGACTCTGAAGAAGAAAAGAGCTTTATTTAAACCTAGCT 250
181 TTACATCTGCTGATGATTTTCTCTCAAGCATCTGCTTAAGAGTTCGCTTACTGCTA 240
|||||
251 TTACATCTGCTGATGATTTTCTCTCAAGCATCTGCTTAAGAGTTCGCTTACTGCTA 310
241 GCCGTGCTGCTGCTGATATTTTTCAGATTTATGCTCTGAAGCTCCTTACACATCCCT 300
|||||
311 GCCGTGCTGCTGCTGATATTTTTCAGATTTATGCTCTGAAGCTCCTTACACATCCCT 370
301 GATTAACATAAGGATATATTTATGTTTATTAACAAGACAGTTGAAGGGGCTAGAGATACA 360
371 GATTAACATAAGGATATATTTATGTTTATTAACAAGACAGTTGAAGGGGCTAGAGATACA 430
361 AAGAGCCACAAATTCATAGGATTTTATTTACTGTGAGACATGCTTGGGTCAAGTCA 420
|||||
431 AAGAGCCACAAATTCATAGGATTTTATTTACTGTGAGACATGCTTGGGTCAAGTCA 490
421 TATTAACATATGCTTTGAGTTAGAGATAGCAATGAAATTTTACCCAGCTTATACAGAAC 480
|||||
491 TATTAACATATGCTTTGAGTTAGAGATAGCAATGAAATTTTACCCAGCTTATACAGAAC 550
481 TTATTTTACGTTATTAACAATGCGCCACATCAGAAAGTCCATATGACATGTTAGACCTT 540
|||||
551 TTATTTTACGTTATTAACAATGCGCCACATCAGAAAGTCCATATGACATGTTAGACCTT 610

541 ATGACTCTATTTATTTTGAAGGTATATACAGTGTCTCAGAGCTTTTGGATACGTTTAA 600
|||||
611 ATGACTCTATTTATTTTGAAGGTATATACAGTGTCTCAGAGCTTTTGGATACGTTTAA 670
601 GTAATTCGTGATCCGCTCATATGAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 660
|||||
671 GTAATTCGTGATCCGCTCATATGAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 730
661 TTACTGAAGAGACAGCTCAAGCTTATGAGCATATATTTACACACTTTTATATCAGGTT 720
|||||
731 TTACTGAAGAGACAGCTCAAGCTTATGAGCATATATTTACCAATTTTATATCAGGTT 790
721 CTGATGCTTGGGAAACATCTATACAGCATTTTGCAGAGCATGCTTTGACTTAATTTTG 780
791 CTGATGCTTGGGAAACATCTATACAGCATTTTGCAGAGCATGCTTTGACTTAATTTTG 850
781 GAGCTCTCAATATTTGATATGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
851 GAGCTCTCAATATTTGATATGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 910
841 TTAAGAGCAATGATATGAGAGAGCGCTTACAGTTGTTAAACTACTGGCAAAAATGTTT 900
|||||
911 TTAAGAGCAATGATATGAGAGAGCGCTTACAGTTGTTAAACTACTGGCAAAAATGTTT 970
901 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAAACAGCCACTTTGGCAGTCTTGGGC 960
971 GGGGCAAGAGATTCAGAAATTTGGCTTCTCAAAAACAGCCACTTTGGCAGTCTTGGGC 1030
961 AGGTTATGATATTCATATGATATGATATGATATGATATGATATGATATGATATGATAT 1020
1031 AGGTTATGATATTCATATGATATGATATGATATGATATGATATGATATGATATGATAT 1090
1021 CTCATGAACCATCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGTGAAGTACAT 1080
|||||
1091 CTCATGAACCATCTGATTTAGCAAAAGACTTAAACAGATATCTTAAAGTGAAGTACAT 1150
1081 GACCTGAGGAAGCATTTAGACATGATGTTTATGCTCAATATGTTACAGCTGCTTAAAG 1140
1151 GACCTGAGGAAGCATTTAGACATGATGTTTATGCTCAATATGTTACAGCTGCTTAAAG 1210
1141 GATATTTCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
1211 GATATTTCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1270
1201 CGATGAGAGTACGCAAAAGAGCCATGATGAGGACTTGCCTCAAAATTTATTAAGAAATGCT 1260
1271 CGATGAGAGTACGCAAAAGAGCCATGATGAGGACTTGCCTCAAAATTTATTAAGAAATGCT 1330
1261 TTACAGTACAGAGCTGGAAGAGATGCTGCCAAACAGATAGAGATGATGATGATGATGAT 1320
1331 TTACAGTACAGAGCTGGAAGAGATGCTGCCAAACAGATAGAGATGATGATGATGATGAT 1390
1321 CTACATATATATTTATCAAAATAGTATGATGATGATGATGATGATGATGATGATGATGAT 1380
1391 CTACATATATATTTATCAAAATAGTATGATGATGATGATGATGATGATGATGATGATGAT 1450
1381 CAATACATGTTTCTCACAATTTTGAAGCTACAGAGAGGATGAAGTCTTATTTACTTGG 1440
1451 CAATACATGTTTCTCACAATTTTGAAGCTACAGAGAGGATGAAGTCTTATTTACTTGG 1510
1441 TATGCAACACATGATTTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1511 TATGCAACACATGATTTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1568
1501 CTGCTCGACATCAAGTAAAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1569 TTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1628
1561 AGTGTCAAGGCCATATTTTCAAAAGTAT 1589
1629 AATGTAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1657


```

RESULT 9
AB014548      5177 bp  mRNA      linear  PRI 06-FEB-1999
LOCUS      Homo sapiens mRNA for KIAA0648 protein, partial cds.
DEFINITION
ACCESSION  AB014548
VERSION     AB014548.1  GI:3327109
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (sites)
AUTHORS     Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
            Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
JOURNAL     DNA Res. 5 (3), 169-176 (1998)
MEDLINE     98403880
PUBMED      9734811
REFERENCE   2 (bases 1 to 5177)
AUTHORS     Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
            Direct Submission
            Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdna1info@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES
    source
        1..5177
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="HJ03994"
                /sex="male"
                /tissue_type="brain"
                /clone_lib="PbluescriptII SK plus"
                /dev_stage="adult"
                1..5177
                /gene="KIAA0648"
                <1..2558
                /gene="KIAA0648"
                /codon_start=3
                /product="KIAA0648 protein"
                /protein_id="BAA31623.1"
                /db_xref="GI:3327110"
                /translation="IYYIVASLDPNVAKALINEMKQNMRLSHVRELLDIHKQPTSEA
                MCSAMGKIMTAKNLPDRKQKQDVPYKKKQVYGDDEKRLSLELLISPTSCKQADI
                CYREIARKLANEKQPTNPFLKMYKFLERLAPVHIDSEALVAKLANKSIKSTADE
                EGEVSPDTAIRSGLELLKLVSTFHPFSFAETYESLQELRMEDDKVAALAIQIFEN
                TGHKIDLPQIRSLIPILHOKAKGTPHQAQVACHAIFTNKQVOLAQIIEPNS
                RSLNADVPQOLITPLVLSLHISMLAPDQSPKMSVAVNITVDLMLNDSGTEKNS
                LMSPEDEVPEVLAKVOAKILYIRMLGMKNOSKNSSTLRLLSLAMLVESGLTEBK
                RISKSMRRLAAGSAILKIMLQEPCHETITPEOFOCALYINDCYQVROIFAKL
                HKALVLLPLELMAIFALCAKDPYKERRAHARQCLKNTSIRETIKQPMATEKL
                SLIPYVVPYMIHLAHDPDTRSDVDLADIKELFMLEVLMTKNENNSHAFMK
                MAENIKLRDAOSPDSEKTNKELTYTCDVALCYINSKSLCNDSPDPLPKMFTFO
                PEKDFNDKSYISEETRVLLITGPKPKAGVAVNKPLSATGKPYRSGTETSGSI
                NVNSELNPSGTSGRSROSSAETGVSSENVENPRIISVPKNIIDPKKKEKSLNSDA
                TQGNISDGGKKRRTYVAAAGENTQCKTDEKVDVDSGPPAPSKPRGRPRKRSQNMATK
                NDDLKRPINKRRRAVVGESPEGLGNNKAKFKLDLAKKAPAEQDIDLQK
                ORIGIN
                BASE COUNT      1690 a      931 c      1037 g      1519 t
                Query Match      25.0%; Score 1041.4; DB 9; Length 5177;
                Best Local Similarity 68.0%; Pred. No. 3.5e-204;
                Matches 1450; Conservative 0; Mismatches 681; Indels 0; Gaps 0;
    QY      1427 GCTTATATCTGATGACCACTGGATTTAAATGCTGTGAAGCATTTGAATGAATGT 1486
            ||||||| ||||| ||||| ||||||| ||||| ||||| ||||| |||||
    Db       1 GCTTATATCTGATATATGCTAGTTGGATCCCAATGCTGTAAAGCTCTCAAGCAATGT 60
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1487 GGAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTGATTTAAGCAAC 1546
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

    Db       61 GGAATGTCAAGACATGCTTCGGAGCATGTACGCCACATGATGATTTGACACACAGC 120
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1547 CCAAAACAGATGTCAGATGTCACAAAGGCATATTTTCCAAAAGATGATTTATTAAGAAATTT 1606
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       121 CTACATCAAGAGGCTTAACTGTTCTGTCATGTTTGGAATAAGATGATGACCATTAAGCAAGAAATTT 180
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1607 TACCTGATCTGTGTAAGGCTCAGAGATTTTCATGAAGAAATTCACACAGAGTGTGAAGATG 1666
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       181 TGCTGACCCCGGGAAAGCAGACAGATTTTGTGAAGAAATTTAACAGAGTCTCGGGGATG 240
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1667 ATGAGAAATTAAGAAAGCACTGTGAAGTACTTTGTAGTCCAAATGCTCTGCAACAGC 1726
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       241 ATGAGAAACTTCGGTCTCAGCTGTGAGATTAATTAATTAAGCCCAACCTGCTTTCGAACACAG 300
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1727 CTGAGGTTGTGCGGAAATTAAGTAAAGTGGGACCCCAACACCTTCAAAATC 1786
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       301 CAGATATTTTGTGAGAGAAATATGCCCCGAAACTTGCAAATCCTTAAGCAACCAAAATC 360
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1787 CTTTCTGGAATATGATCAAGTCTTCTGTGAGAGATAGCACCTGTGCACATGATATCCG 1846
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       361 CTTTCTGAGATATGTCAAATTTCTGTGGAAGAAATGCGACCTGTGCACATGATATCCG 420
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1847 AATCTATCAGTCTCTTATTATTAACCAAGTACAAATCAATGATGGAACGCGATGATG 1906
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       421 AACCCATTAAGTGCATAGTCAATGATGATTAAGTCAATGAGGGGACAGCATGATG 480
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1907 AAGATGAGGTTGTCACATGATCAAGCCATCAGAGAGGTCTTGAACCTTAAGGTAC 1966
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       481 AAGAGGAGGTTGTAAGTCCAGATCAAGCTATACGTTGAGAGCTTGAACCTTAAGGTTC 540
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      1967 TCTCATTTTACATATCCATCTCATTTTCTGCTGCAAACTTTGATCATTTACTGACTT 2026
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       541 TGCTTTTACATATCTGACTCTGTTCCAGTCTGACAGCATATGATGCTCTGTTACAGT 600
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2027 GTCTGAAATGATGATGAAAGATGACAGACAGCTGACATGCAATTTTCAAAACACAG 2086
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       601 GCCTAAGAAATGAGGATGACAAAGGTACCAACACTGATTAATTTTGAATAACAG 660
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2087 GAAGCAAAATTTGAAGAGATTTTCCACATGATGATGAGGCTGCTGCTGTTTACATC 2146
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       661 GTCACAAATATGAAACAGACACTTCCCGACATGATGATGATGATGATGATGATGATGATG 720
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2147 ACAAAATTAAGAAAGACCCCGGTCACCAACCAATATGCCATTCATTTATCATATGCA 2206
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       721 AAAAAGCAAGAGGGGTACTCCACACCAACCAAGAGCTGTGACGTATACAGCCCA 780
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2207 TATTTTCTAATAAGAGACCCAGTGTGCAAGATATTTGACCTCTGCAATAAAGCCTAG 2266
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       781 TATTCACAAATTAAGAAAGTCCACGCTGACAGATTTTGTGACCACTAGTGAAGTCTGA 840
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2267 ATCCAAAGCACTGGAACATCTGTATAACACATGTTGATATTTGATGATGATGATGATGATG 2326
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       841 ATGCTGATGTCGACAGAAACATTAATATCTCATATAGTTTATTTGGGCAATTTTATATG 900
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2327 TTGCACCTGATCAATTTGCTGCTCTTGTGAATCTTGGTATGCTACTTTCATTTGTAAG 2386
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       901 TACACACAGATGATGTTGCTTCCCACTGAAATCTGTAGTGAAGCAAAATTTATTTGTGAAG 960
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2387 ATCTTCATGATGATGATGCGCTTCCAGAGGAAAAAGACACTAAACTTTGGTTCAGATG 2446
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       961 ATCTGCAATGAATGACAGAGGTCAACAGTGAAGAAATGGAAGAAACGCTGCTCCAGATG 1020
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2447 AAGAGATATCTCTGACAGCAATGATGTCAAATTAAGGCTATTTAAATGATGATGATGATGATG 2506
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       1021 AAGAGATTTCCCTGAGATGACTAGCAAAAGTACAGCAATTTAAACTCTGTGAAGGTGCG 1080
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2507 TACTTGAATGAAAAATATATCACAGTAATCAGGAACCTTCTACCTTAAGATGCTACAA 2566
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    Db       1081 TGTTGGTATGAAAAACACACAGCTTAATCTGCCAATTCACACCTTGGTATTTATGAC 1140
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    QY      2567 CAATATGCTAGTATGATGAGACTTGACAGAAAGGGAATTTAGTAAACAGATATGT 2626
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 1141 CGATGTGGTAGTGAGGGGTGACCTGACAGCAAAAGAGATCAATGATATGT 1200
 QY 2627 CAGCTGTGAGCTGCTGCTGGAGTGCATATGTGTAAGCTGCACAAAGACCTGTACC 2686
 Db 1201 CTCGGCTGGATTAGCTGCTGGTAGTGCCATATGTAAGCTGCTCAGGAACCTGTGTACC 1260
 QY 2687 ATGAAATCATCATATTAGAAACAATATACATATGTGCTATGCTATCAAGCAATGCT 2746
 Db 1261 ATGAAATTTATACCCAGAAAGATTTCACCTCTGTGCACCTGTGTATATGATGAGTGT 1320
 QY 2747 ATCAAGTAAGCAAGTGTGGCCAGAAACCTTACAAAGCCCTTCCGTTACGGCTTC 2806
 Db 1321 ACCAAGTAAGCAGATATTTGCTCAGAACCTCATAGCAGATATGGAAGTTACGTCC 1380
 QY 2807 CACTGAGTATATGCAATCTGCTGCTTGTGCAAAAGATCTGTAAAGAGAGAAAG 2866
 Db 1381 CATTGGAGTATATGGGATCTTGTGCTGTGTGCAAAATCTCTGGAAGGAGAGAAAG 1440
 QY 2867 CTCATGCTAAGCAATGTTTGTGTAATAATATGTAAGCGCGAGATCTGGAAGCAGC 2926
 Db 1441 CACAGCAGCAACAATGTTTACTGAAATAATATCAGATATACGAGGGAATATTAAGCAGA 1500
 QY 2927 ATGCAAGCTGTAGTAAAAATATATGTCTCTCTACAGAGATATGTTGTCATATACA 2986
 Db 1501 ATCTATGGCTTACTAGAAATATATATCATCTGTGCTGGAATATGATGTATCATATCA 1560
 QY 2987 TTCACCTTTTGGCAGATGACCCAGATTTATGTCAAGTACAGATATGTAACAACCTTAAAG 3046
 Db 1561 TTCACCTGCTAGCCCATGATTCAGATTTTACAAGATCACAGATGTGTGACGCTTCGTG 1620
 QY 3047 ATGTTAAGAAATGTCTTGTGTTTGTGGAATATTAATGCTAAAAATGAAATATACA 3106
 Db 1621 ATATCAAGAGTGCCTTATGCTTATGAGTTTATATGACAAAGATGAAACAAATTA 1680
 QY 3107 GTACAGCTTTTATGCAAAAGATGTAGAAATATTTAACAACAAAGATGCCCAAGCAGC 3166
 Db 1681 GCCAAGCTTTATGTAAGAAAGATGCAAGAACATCAAGTTTATCCAGAGATGCCAGTCTC 1740
 QY 3167 CAGATGATGCAAAAAATGATGAAACATGTACATGTGTGATGTGTGCATGAAATATCA 3226
 Db 1741 CAGATGATCCAAAGCAAAATGAAACATGTATATACATGTGTGTGTGTCTCTGTGTTA 1800
 QY 3227 TCATGTCAAAGATGCTACTACATACAGTTTGAATCTCTAAAGACCCGCTACTACAGCTC 3286
 Db 1801 TAAATAGTAAAGTCTTGTGTCATGAGATTCACAAAGAGACCCAGTCCCAATGA 1860
 QY 3287 GTTCTTTCATCACTGACACAAAGATTTTCAGTAACACCAAAATTTCTGCTCCTGA 3346
 Db 1861 AATTTTATACAAACCTTAAAGAGATCTCTGTAAGATTAAGATTATATTTTACAAGAGA 1920
 QY 3347 TGAATTCATTTTCTCTGCTGGAACCTTAAACCAACCAATGTTAGAGAGCTGTAAACA 3406
 Db 1921 CAAGAGTACTTCTGTTAAAGAGAAAGCCCAAGACCTGCTGAGTACTAGTGTGCAATTA 1980
 QY 3407 AGCCATTTTCATCAGCAGCAGCAAGCAATCTCAGACCAAAATCATCAGAAATGTA 3466
 Db 1981 AGCCTTTATCAGCAAGGAGGAAGAACCTTATGTAAGACCTGCACTGAGACTGGAA 2040
 QY 3467 GCAATGCAAGCAGCAGCTCAAAATTCAGAGCTCTCTGGAAGAAATTAAGGAGAGCTTGATA 3526
 Db 2041 GCAAATTTAATGTAATTCAGAGCTGAACCCCTTCAACCGGAATGATCAACGGAACAGA 2100
 QY 3527 GTTCTGAATGATCAGAGTGAATGAAGA 3557
 Db 2101 GTTCAGAGCAGCAGAAACTGGAGTTAGTGA 2131

RESULT 10
 LOCUS BC021408 1976 bp mRNA linear ROD 16-APR-2003
 DEFINITION Mus musculus RIKEN cDNA 9030416h16 gene, mRNA (cDNA clone MGC:29439
 IMAGE:3964501), complete cds.
 ACCESSION BC021408

VERSION BC021408.1 GI:1820409
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 1976)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheet, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shewchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22368257
 12477932
 2 (bases 1 to 1976)
 Strausberg, R.
 Direct Submission
 Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT: MGC help desk
 Email: cgabds-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nrl.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., LaRic, P., Legspt, R., Maduro, O.L., Mastello, C., Maskell, B., Mastrian, S.D., McKloskey, J.C., McDowell, J., Pearson, R., Stantip, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 39 Row: b Column: 24
 This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.
 Location/Qualifiers
 1..1976
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:29439 IMAGE:3964501"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI_CGAP_Mam1"

QY 3208 GATGTTGCCATGATATCATGTCAGAGACTACATACAGTTGGAACTCCTTAA 3267
11
Db 1765 GACGTGCTCTGTGTATTAATAAGTCTTGTGCATGCAGATCACCAGA 1824
QY 3268 GACCCGGTACTACACCTGTTCTTCTACTCAACCTGACAGATTTCAGTACCA 3327
11
Db 1825 GACCCAGTCTCCCATGAGATGTTCTTACACAGCCGAAAGGTAATTCCTCACA 1884
QY 3328 AATTAATCTGCCCTCCGAATGAATCATTTTCAC 3363
11
Db 1885 TTTCCTGTCTATGTAAGTAATGATTAATTCAGT 1920

RESULT 11
AK098331
LOCUS
DEFINITION
AK098331 2265 bp mRNA linear PRI 15-JUL-2002
Homo sapiens CDNA FLJ41012 fis, clone UTR02018523, highly similar
to Human androgen-induced prostate proliferative shut-off associated
protein (AS3) mRNA.
AK098331
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens (human)
oligo cloning: fis (full insert sequence).
Eukaryota; Eutheria; Primates; Catarrhini; Homidae; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS
1 Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Ito, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, Y., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Magatsu, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K.
and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2265)
Isogai, T. and Yamamoto, T.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5- & 3-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
Source
location/Qualifiers
1..2265
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UTR02018523"
/tissue_type="uterus"
/clone_lib="UTR02"
/note="cloning vector: pME18SF13"
125..1927
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC05286.1"
/db_xref="GI:21758324"
/translation="MDTAPQRPATALCGVVSADGIAPPGVKEITDKITDEMIR
KMWVTFMDHDOSEDEKQOYLPLALHLASFLELRNPKDVRLLVACCLADIFPIA
PEAPYSHDKLDFLFTROLKGLDTSKQFNRYFYLLLEMLAWKSYNCELEDC
NEIFQLPRTLESVINSHNKKVOMHMDLMSIIMEGVTYOEELSLNLPANH
NLNOSPLALVILKRTVOTTEACIANFNQYLVGRSSVSLDSHVVLDIOELAFD
PHILSYNPOLPEFKLSNDEBERLAVVLLKLRSKSDSLTORPILMOCTLEAFD
IHVPRLSVKFAHCLMHPDLADLLEYLKVSRSHDEALRHQVITLITTAARLD
ALVNDQLLSFVRETRIDKRWVRKEMAMGLQLYKRYLHGAGKEAEKSWIMDKL
LHIYONSIDKLIVERKIFPAQYLVPHNLETBRMKLYLVASLDPNVAKLNEWMKC

BASE COUNT 696 a 427 c 476 g 666 t
ORIGIN
Query Match 22.3%; Score 932.2; DB 9; Length 2265;
Best Local Similarity 71.5%; Pred. No. 1.3e-181;
Matches 1225; Conservative 0; Mismatches 488; Indels 0; Gaps 0;
ONMLSRVRELDLHKOPTSEANCSAMEFKLMTIAKNLPDPKADPFVKFENQVLGDD
EKLRSOLELLISPTCSKQADICVYKSFYFLF"
QY 28 GATGAAAATTTACATATCCGCTGGGCTCAAGAAATVYCATAAATATCTAAAGAG 87
11
Db 182 GAGGGAAATGCTGTTACCTCCGGGGTAAAGAGATCACCAGATCACCAGAC 241
QY 88 GAGATGCTGAGCAATTAAGATGCTTGAAGAACTTTATGATATGACACGAGCT 147
11
Db 242 GAGATGATCAACGCTCGAAGATGAGTGAAGAACTTTATGATATGATGATGAGCTGA 301
QY 148 GAAGAAGAAAGAGAGCTTATTTAAACCTAGCTTACATCTGCTTCAAGATTTTCTC 207
11
Db 302 GAAGATGAAAACAGCAGTATCTCCGACATAGCCTTGATCTGATCTGAATCTCTC 361
QY 208 AAGCATCTGTGAAGATGCTTCTGATGAGCTGCTGCTGCTGATATTTTCAAG 267
11
Db 362 AGGAACCCCAATAAAGATGCTGCTCTGTAGCATGTTGCTGATATCTTCTGCT 421
QY 268 ATTTATGCTCTGAAAGCTCTTACACATCCCTGATAACTAAGATATATTTATGTT 327
11
Db 422 ATCTATGCCCCAGAGCTCTCATATCTTCCATGATAACTTAAGACATATTTTGT 481
QY 328 ATACAGACAGCTTGAAGGGCTAGAGATACAAAGGCCCAATTCATAGTATTTT 387
11
Db 482 ATTACACAGCATTTAAAGCTTTGGAGATACAAAGATGACAGTTTAATGATCTT 541
QY 388 TATTTACTGACACATCTGCTTGGCTCAAGTATATATATATGCTTTGATAGAT 447
11
Db 542 TATTTATGAGAAATTTAGCTTGGTGTAAATCATATACATCTGCTTGAATGGAAT 601
QY 448 AGCAATGAATTTTACCCAGCTATACACATCTTACGTTATTAACAAATGGCCAC 507
11
Db 602 TGCATGAATTTTATTTATGCTTTTAAACCTCTCTTCAAGTATCAACAAATGGCCAC 661
QY 508 AATCAGAAAGTCCATATGACATGATGATGATGATGATGATGATGATGATGATGAT 567
11
Db 662 AATGAAGAGTACAAATGACATGATGATGATGATGATGATGATGATGATGATGAT 721
QY 568 ACAGTCTCAGAGACCTTTGGATACGTTTATGATATCTGATCTGATCAAGAAT 627
11
Db 722 GAGTGTACTCAAGAAATTTATGATGATGATGATGATGATGATGATGATGATGAT 781
QY 628 TTAACCAAGCAAGCATATGATTTTGGCAAGCTTTACGAAAGAGACGCTCAAGCTAT 687
11
Db 782 TTAATTAACCAAGCTTTGACCTTCAAAAGGCTATTAAGAAAGCAAGCTCAAGCTAT 841
QY 688 GAGCATATATATACACCTTTTATATGATGATGATGATGATGATGATGATGATGAT 747
11
Db 842 GAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
QY 748 GATTTGTGACAGCATGCTTTTGAATTTTGAAGCTTCAATATGATGATGATGAT 807
11
Db 902 GATTTGTGACAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 961
QY 808 CTGCTCTGCTTTTACCCAGCTTGAATTTTAAAGAGCAAGATATGATGATGATGAT 867
11
Db 962 TTAATTAACCAAGCTTTGACCTTCAAAAGGCTATTAAGAAAGCAAGCTCAAGCTAT 1021
QY 868 CTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 927
11
Db 1022 TTAGTGTGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081
QY 928 CAAGAAGCAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 987
11
Db 1082 CAGATGCTCTCTTGGCAAGTATTTCTTGGACATTTAATGATATTCATGCTCTGTG 1141

QY	988	CGCCTGGAATGTGGAAATTTGCAAGCATTGCTCTCTGAACCCCTCTGATTTAGCAAA	1047
Db	1142	AGATTAAAGATGGAAATTTGGCCAGTCATTTGTTAATGATATCCACAGATTTTAGCBAAG	1201
QY	1048	GACCTTAACAGAGATATCTTTAAAGTGAGGCTCACAATGACCCCTGAGAAAGCTATTAGACAGAT	1107
Db	1202	GATCTCAACAGAAATTTTAAAGGTTAAGATACATGATATCCAGAAAGAAAGCTATTTCGTCATGAT	1261
QY	1108	GTTATTTGTGCATATAGTTTACAGCTGCTTAAAAAGGATATTTCTTTCGTCATGATCACTTA	1167
Db	1262	GTCATTTGTCTACTAATAATACACAGCTGCCAGAGGAGCCCTGGCTTATGTAATATGATCACACTG	1321
QY	1168	CTTAATTTTGTGACAGAGAGAACATTTGACAAAGATGGAAGAGATACCCAAAGAGCATG	1227
Db	1322	CTTGCGCTTTGTGAAGGAAAGACACTGATTAACGATGGCGCAGTAAGAAAAAGAGCATG	1381
QY	1228	ATGGGACTTGGCCCAAAATTTATAGAAATATGCTTTTACAGTCAGCAGCGCTGCAAAAGATCT	1287
Db	1382	ATGGGCTCGGCTCAGCTTTATTAAGAAATACGTCTTCATGCTGAAGCAGGAAAGGAAGCT	1441
QY	1288	GCAAAACAGATAGCATGGATCAAAAGACAAATTCGTACATATATATTAACAAATAGATTT	1347
Db	1442	GCAGAGAAAGTCACCTGATTAAGGACAAACTTCGATATTTATTTATCAGAACGACATT	1501
QY	1348	GATGATGAGCTACTTGTGGAAGGAGATCTTGTCCTCAATACATGCTCTCCCAATTTTGA	1407
Db	1502	GACGACAAACTGTTGGTAGAGAAAATCTTTGGCTGATGATCTTGCCCCCAACCTTGAA	1561
QY	1408	ACTACAGAACGGATGAATGCTTATATCTTGTATTCGCAACACTGAGATTTAAATGCTGTG	1467
Db	1562	ACAGACAGAGAAATGAATGCTTATATCTTATATGCTATGCTATGTTGGATCCAAATGCTGA	1621
QY	1468	AAAGCATTTGAATGAATGTGGAATGTCAAATCTGCTCCGACATCAAGTAAGAAGATTTG	1527
Db	1622	AAAGCTCTCAACGAAATGTGGAAAGTGCAGAACATGCTTGGAGCACAATGTCGCGAACTA	1681
QY	1528	CTTACCTTGATTAAGCAACCCAAACAAAGATGCCAGTGCAAAGGCCATATTTTCCAAAGTG	1587
Db	1682	TTGGATTTGCCAAAGCAGCCCTACATCAAGGCTTACTGTTTCTGCGCATGTTTGGAAAACTG	1741
QY	1588	ATGGTTATTACAGAAATTTTACCTGATCTCTGGTAAGGCTCAGATTTTCATGAAAGAAATTC	1647
Db	1742	ATGACCAATAGCAAAAGAAATTTTCCTGACCCCGGGAAGACAAAGATTTTGTGAAGAAATTT	1801
QY	1648	ACACAGGTGTTAGAAGATGATGAGAAAAATTAAGAAAGCATTTAGAATATCTTGTATGCTCA	1707
Db	1802	AACCAAGGTTCTCGCGCAGATGATGAGAAAACTTCGCTCAGTTGAGGATATTTAATTAGCCCA	1861
QY	1708	ACATGCTCTGCAAGAGGCTGAAGGTTGTGTG	1740
Db	1862	ACCTGTTCTTGCAACACAGATATTTGTGTG	1894

RESULT	12
LOCUS	BC041361
DEFINITION	BC041361 2705 bp mRNA linear PRI 09-JAN-2003
ACCESSION	Homo sapiens, similar to KIAA0648 protein, clone MGC:43818
VERSION	IMAGE:5273075, mRNA, complete cds.
KEYWORDS	BC041361 BC041361.1 GI:27552793
SOURCE	MGC.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL	1 (bases 1 to 2705)
	Strausberg,R.
	Direct Submission
	Submitted (16-DEC-2002) National Institutes of Health, Mammalian
	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT

Contact: MGC help desk
Email: cgapbs-retail.nrh.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Kan Guin,
Letlicia Hsiao, Martin Krzywinski, Reta Kutische, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Nees, Pawan Pandoh, Anna-Lilias Prabhu, Parvaneh Saeedi, Jacqueline
Schelch, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stolt,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>
Series: IRAC Plate: 75 Row: c Column: 19.
Location/Qualifiers
1..2705

FEATURES

SOURCE

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:43818 IMAGE:5273075"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
543..2347
/codon_start=1
/product="Similar to KIAA0648 protein"
/protein_id="AAHA1361.1"
/db_xref="gi:27552794"

CDS

//translation="MDFTAPKRPATFALCGVNSADGTIAYPGVEIKITDTIMIRK
LKWKYFKPMKDSDSEDKQCYDLIAHLISEPFLRPNPDRLVAGCADIARIVRA
PEAVYTSHDKIDKIIFLPTRRLKLEDETKSQSPFRYYLLENLAWSYNICPELEDC
NEITIQLRILFSVINSHSKVKVMHMLDMSILIMGDEVOELSLISLINLPARK
LNMAOSFDLAFVLKRTVOTIEACIANFQNVILGRSVSDISEHPWLDLOEFLAID
PHLSVSFOLEFLIKSNDEGERLAVRVLAFLAKSGSDSLATNPRLWOCFLERFND
IHVVRLSEVVFASHGLMNHEDLAKLDLEYLKVRSHPBEKAIRDVIYVITTAARDL
ALVNDOLGFYERETDRMRVRKEAMGAOLKKYCLHGEAEKVAEKVSTKDKL
LIHYONSIDKLKIPKIFROYLYPHULEFERPKCLYLYASIDPVNAVLANMMCK
ONMRSHVRELLDLHKOPTSBAKSNAPFKMTAKNLPDGKADQEVKKFNQVLGDSD
EKHSOLELISPSCSKQADICVASSYFFLF"

BASE COUNT

751 a 589 c 654 g 711 t

ORIGIN

Query Match 22.3% Score 932.2 DB 9 Length 2705;
Best Local Similarity 71.5% Pred. No. 1.3e-181;
Matches 1225; Conservative 0; Mismatches 488; Indels 0; Gaps 0;

28 GATGGAAAAATTACATATCCGCTGGGGTCAGAAGAAATATCAGATAAATATCTAAAGAG 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 602 GACGGGAAGATCGCTTACCCTCCGCGGGTAAGAGATCACCGACAAGATCACCAAGGAC 661
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

88 GAGATGCTGACACGATTAAAGATGTTTGAAAACCTTTATNGATATGACCAAGACTCT 147
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 662 GAGATGATCAACACGCCCTGMAAGATGGTAGTGAACACCTTATAGATNGATGACGACCA 721
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

148 GAAGAAGAAAAGAGCTTATTATTAACCTTACCTTACCTTACCTTACCTTACCTTACCTT 207
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 722 GAAGATGAAAACACGACAGATATCTCCCATAGCTTGACCTTGATCTGAAATTTCTCTC 781
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

208 AAGCATCTGTTAAAGATGTTGCTTACTGTTAGTACCTGCTGCTGCTGATATTTTCAGG 267
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 782 AGGAACCCCAATATAAGATGTGCGTCTCTCTTATAGATNTGTTGGCTGATATCTTCGT 841
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

268 ATTATGCTCTGGAAGCTCTTACACATCCCTCATATAACTAAAGATATATTATGTTT 327
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 842 ATCTATGCCCCAGAGCTCCATATACCTTCATATTAAGGATTAAGGATTTTTTTGTTT 901

QY 328 ATACAGACAGTTGAAGGGCTAGAGATACAAAGGCCAACATTCATATGATATTT 387
 11
 Db 902 ATTACAGACATTAAGAGTTTGGAGATACAAAGAGTCCACAGTTTATATGATCTTT 961
 QY 388 TATTTACTGAGAACATTCCTGGGTCAAGTCAATATACATATGCTTTGATTAAGAT 447
 11
 Db 962 TATTTATAGAAATTTACTTGGGTAAATCATATATACATCTGCTTAAATGGAAGAT 1021
 QY 448 AACAAATGAATTTTCCAGCATATGACACTTATTTTCAATTTTAAACATGGCCAC 507
 11
 Db 1022 TCCAAATGAATTTTATTCAGCTTTTAAACCTCTCTTCAGATGATCAACATATCCAC 1081
 QY 508 AATCAGAAAGTCCATATGACATGATGACCTATAGCTATATATTTTGTGAAGTGTAT 567
 11
 Db 1082 AATTAAGAGGTACAAATGACATGATGATGATGATGATGATGATGATGATGATGAT 1141
 QY 568 ACAGTGTCTCAGAGAGCTTTTGAATGCGTTTAAATGATGATGATGATGATGATGAT 627
 11
 Db 1142 GGAGTACTCAGAAATATTTGGACCTCATTTATTTACCTCATTTCTGACATTAAGAC 1201
 QY 628 TTAACCAAGCAAGCATATGATTTGGCAAGGCTTTCTGAAGGACAGCTCAAGCTATT 687
 11
 Db 1202 TTAATTAACAGCTCTTTGACCTGCAAAAGTGTATTAAGAAAGACAGTCCAGATATT 1261
 QY 688 GAGCATATATTTACACCTTTTATTAATCAGTTCATGATCTTGGAAACATCTATCAGC 747
 11
 Db 1262 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1321
 QY 748 GATTTGTACAGCATGCTTTTGAATTTTGGAGCTGTACATATTTGATATGATCTATT 807
 11
 Db 1322 GATTTGTACAGCATGCTTTTGAATTTTGGAGCTGTATGATGATGATGATGATGAT 1381
 QY 808 CAGCTCTGCTTTTACCAGCTGTGAATTTAAATTAAGCAATATATGAGAGCCG 867
 11
 Db 1382 TTAATTAACAGCTCTTTGACCTGCAAAAGTGTATTAAGAAAGACAGTCCAGATATT 1441
 QY 868 CTACAGATTTGTTAACTTGGCAAAATTTGTTGGGCAAGAGTCAAGATTTGCTTCT 927
 11
 Db 1442 TTAGCTGTTGTTGCACTTCTAGCTAAATTTGTTGGTCCAAAGATTTCTGATTTGGCAACA 1501
 QY 928 CAAACACAGCCACTTTGGCAGTGTCTATTTGGCAGGTTTAATGATATTCATGTCAATC 987
 11
 Db 1502 CAGAAATCGCTCTTGGCAATGTTTCTTGGACGATTTAATGATATTTGTTCTCTGTG 1561
 QY 988 GCGCTGATGTTGTTGCTAGCCATTTGCTCATATACCATTCATTTAGCAAAA 1047
 11
 Db 1562 AGATTAAGAAAGTGTGAATTTGTCAGTCAATTTGTTAATTAATCAACCAATTTAGCGAAG 1621
 QY 1048 GACTTAACAGACTATCTTAAAGTGAAGTGCATGACCTGAGGAAGCTTTAGACATGTAT 1107
 11
 Db 1622 GATTCACAGAAATTTAAAGTTAGATTCATGATCCAGGAAGACTTTGCTCATGAT 1681
 QY 1108 GTTATTTGTCAATAGTTACAGCTGTAAAGATATTTCTTGTGTCATATGATCACTTA 1167
 11
 Db 1682 GTCAATTTGTCTATTAATTAACCTGTCAGAGAGGACTGCTTGTATTAATGATGACGTG 1741
 QY 1168 CTTAATTTTGTGAGAGAGAACATTTAGCAAAAGCATGAGAGTACGCAAAAGAACGATG 1227
 11
 Db 1742 CTTCGCTTTGTAAAGGAAAGAACACTGTAAACGGTGCAGTGAAGAAAGAGCTAAG 1801
 QY 1228 ATGGGACTTTGCCCAATTTTAAAGAAATTTGCTTTACAGTCAAGAGCTGAAAAGATGCT 1287
 11
 Db 1802 ATGGGCTGCTGCTCAGCTTTATTAAGAAATTTCTGCTTCATGATGAGAACAGGAAGAGCT 1861
 QY 1288 GCAAAACAGATAGCATGATGATCAAGCAAAATTTGCTATATATATTTATCAAAATGATATT 1347
 11
 Db 1862 GCAGGAAAGTACGTGATTAAGAGCAAACTTCTGCAATTTATATATCAGAAACAGCATT 1921
 QY 1348 GATGATGACATCTTGTGTAAGAGATTTGCTCATATACATGATGTTCTCACAATTTAGAA 1407
 11
 Db 1922 GACGCAAACTGTTGTGTAAGAAATTTTGTCTCATATCTTGTCTCCCAACACTGGA 1981

QY 1408 ACTACGAACGATGAATTCCTTATATTTACTTGTATGCCACACTGGATTTAAATGCTGTG 1467
 11
 Db 1982 ACAGAAAGAGAGATGAATGCTTATATTTACTATATGCTAGTTTGGATCCAAATGCTGTA 2041
 QY 1468 AAAGCATTTGAATGAATTTGGAAATTTGCAAAATTTCTCTCCGACATCAATTAAGATTTTG 1527
 11
 Db 2042 AAAGCTCTCAAGAAATTTGGAAAGTGTGGAACATCTTCCGAGAGCATTCAGCGCAACTA 2101
 QY 1528 CTTCAGTTGATTTAAGCAACCCAAACAGATGCGAGTGCAGGCTATTTTCAAAAGT 1587
 11
 Db 2102 TTGGATTTTACAGAGCAGCTTACATGAGAGCTTACTGTTTCGCCATTTTGGAAACAGT 2161
 QY 1588 ATGTTTATTAAGAAATTTTACCTGATCTCTGTTAAGCTCAGAGATTTTCAAGAAATTC 1647
 11
 Db 2162 ATGACATATGCAAGAAATTTTGGCTGACCCCGGAAAGCAAGATTTTGTGAAGAAATTT 2221
 QY 1648 ACACAGCTTTTACAGATGATGAGAAATTAAGAAAGCACTTGAAGTACTTGTATGCA 1707
 11
 Db 2222 AACAGGTTCTCGGCATGATGAGAAACTTCGCTCAGTTGGAGTTATTAATTTAGCCCA 2281
 QY 1708 ACATGCTCTGCAAGCAGCTGAGGTTGTTG 1740
 11
 Db 2282 ACCTGTTCTTGCACACAGCAGATATTGTGTG 2314

RESULT 13
 AK026889 2079 bp mRNA linear PRI 29-SEP-2000
 LOCUS AK026889 Homo sapiens cDNA: FLJ23236 fls, clone COL00725.
 DEFINITION AK026889 AK026889.1 GI:10439854
 ACCESSION AK026889.1 GI:10439854
 VERSION AK026889.1
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 AUTHORS Kawabata, A., Hkiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,
 Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
 Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
 MEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 2079)
 SUGANO, S., SUZUKI, Y., Ota, T., Obayashi, M., Nishi, T.,
 Shibahara, T., Tanaka, T. and Nakamura, Y.

TITLE
 JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure Analysis, Human
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
 Japan (E-mail: cdna1@hgc.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
 Fax: 81-3-5449-5416)

COMMENT MEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing; Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES
 source location/Qualifiers
 1..2079

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="COL00725"
 /issue_type="COL"
 /clone_lib="COL"
 /note="cloning vector pME18SFL3"
 304..681
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB15584.1"
 /db_xref="GI:10439854"
 /translation="MDLRLVQEQPKGSORRRKRGHTASESDQWPEERLKEIDI

Db 1922 GACGCAAACTGTTGTGTAAGAAATTTTGTCTCATATCTTGTCTCCCAACACTGGA 1981

CDS
/evidence=not_experimental
join(<4320..4524,6830..6944,9075..9207,9522..9641,
10616..10678,11563..11708,15477..15582,21108..21547,
21638..21748,24176..24311,26004..26039)
/gene="49J10.1"
/note="continued from 267P19.1 in Em:275889
match: proteins: Tr:Q9Y2I5"
/codon_start=1
/evidence=not_experimental
/product="49J10.1.1 (KIA0979, isoform 1)"
/protein_id="CAC94788.1"
/db_xref="GI:16151354"
/translation="DECYOVQVFAQKLHKGSLRLPLEYMAICALCAKDPYKERRA
HAROCILVKNINVRREYLKQAAVSEKLSLPEYVVPYTHLAHPDVKYVDIOL
KVCECLVNLVLEILMAKNENNSHAFIRKVENIKQKQDAGPDDAKMNETLYVCYA
MNINSKSTTYSLESKPDVLPARFTOPKNSNTKNTLPPEMKSFTPGPKTTNV
LSAVNKPILSSAGKQSQTKSSRMETVSNASSNSPGRKGLDSDSEMDHSENYT
MSSPLPGKSKDRDSDLVRSLELPRGRKKTPTVEQELGMDLTLKVOEQPKGS
ORSRKRGHTASEDEQOQWPEERKLKEDILENEDQNSPPKGGRRPPLGGGTRE
EPTMKTSKSGSKSGPPAPEEEEREROSGNTGEOKSKSGHVRRAQOSRAESPS
SALESQSTPQKGRGRSPKTPSPQKKNVVRGRSQAAATKENDSSEVDYFEGSSPV
DLPQETEEEEESTVNVRRRSKRRER"
join(<4320..4524,6830..6944,9075..9207,9522..9641,
10616..10678,11563..11708,15477..15582,21108..21547,
21641..21752)
/gene="49J10.1"
/note="continued from 267P19.1 in Em:275889
match: proteins: Tr:Q9Y451"
/codon_start=1
/evidence=not_experimental
/product="49J10.1.2 (androgein-induced prostate
proliferative shut-off associated protein, isoform 2)"
/protein_id="CAC94787.1"
/db_xref="GI:16151353"
/translation="DECYOVQVFAQKLHKGSLRLPLEYMAICALCAKDPYKERRA
HAROCILVKNINVRREYLKQAAVSEKLSLPEYVVPYTHLAHPDVKYVDIOL
KVCECLVNLVLEILMAKNENNSHAFIRKVENIKQKQDAGPDDAKMNETLYVCYA
MNINSKSTTYSLESKPDVLPARFTOPKNSNTKNTLPPEMKSFTPGPKTTNV
LSAVNKPILSSAGKQSQTKSSRMETVSNASSNSPGRKGLDSDSEMDHSENYT
MSSPLPGKSKDRDSDLVRSLELPRGRKKTPTVEQELGMDLTLKVOEQPKGS
ORSRKRGHTASEDEQOQWPEERKLKEDILENEDQNSPPKGGRRPPLGGGTRE
EPTMKTSKSGSKSGPPAPEEEEREROSGNTGEOKSKSGHVRRAQOSRAESPS
SALESQSTPQKGRGRSPKTPSPQKKNVVRGRSQAAATKENDSSEVDYFEGSSPV
DLPQETEEEEESTVNVRRRSKRRER"
4959..5316
/note="L2 repeat: matches 1830..2229 of consensus"
5608..5810
/note="MIR repeat: matches 46..262 of consensus"
5874..5953
/note="L2 repeat: matches 2625..2736 of consensus"
6757..6826
/note="35 copies 2 mer tt 80% conserved"
7829..8133
/note="AluX repeat: matches 1..297 of consensus"
8892..9087
/note="98 copies 2 mer tt 59% conserved"
10957..11242
/note="AluY repeat: matches 1..291 of consensus"
12344..13119
/note="L1P16 repeat: matches 5321..6156 of consensus"
14108..14427
/note="L2 repeat: matches 2014..2333 of consensus"
14428..14540
/note="HY1 repeat: matches 1..111 of consensus"
14541..14935
/note="L2 repeat: matches 1619..2014 of consensus"
14951..15154
/note="L2 repeat: matches 2529..2730 of consensus"
16579..16885
/note="AluX repeat: matches 4..312 of consensus"
18059..18397
/note="L1P repeat: matches 3809..4138 of consensus"
19076..19481
/note="MLT1A2 repeat: matches 1..374 of consensus"

repeat_region	19502..19807
/note="AluX repeat: matches 1..310 of consensus"	
repeat_region	19807..20122
/note="AluY repeat: matches 1..294 of consensus"	
repeat_region	23282..23371
/note="L2 repeat: matches 2590..2733 of consensus"	
repeat_region	23734..23944
/note="L2 repeat: matches 1814..2050 of consensus"	
repeat_region	24795..25088
/note="AluX repeat: matches 1..294 of consensus"	
repeat_region	26083..26106
/note="12 copies 2 mer tt 95% conserved"	
misc_feature	26140..26289
/gene="49J10.1"	
/note="match: STS: Em:G43328 Em:G43329"	
26749..27206	
/note="match: GSS: Em:A2019111"	
26790..27245	
/note="match: GSS: Em:A098541"	
26877..26882	
/gene="49J10.1"	
26899	
/gene="49J10.1"	
26900	
/gene="49J10.1"	
26901	
/gene="49J10.1"	
26902	
/gene="49J10.1"	
28327..28360	
/note="17 copies 2 mer tt 88% conserved"	
29194..29299	
/note="match: GSS: Em:A0675002"	
29212..29650	
/note="match: GSS: Em:A0675002"	
29300..29554	
/note="L2 repeat: matches 2498..2750 of consensus"	
Query Match	10.7% Score 447; DB 9; Length 137246;
Best Local Similarity	85.5% Pred. No. 6e-82;
Matches 550; Conservative	0; Mismatches 0; Indels 93; Gaps 1;
QY	3624 GTCGTAATGGAGAACCTTAGAGGACAGAAAAAACGCCGTCACAGAACAGAGAGAGA 3683
DB	21107 GTCGTAATGGAGAACCTTAGAGGACAGAAAAAACGCCGTCACAGAACAGAGAGAGA 21166
QY	3684 ATTAGGTATGATGACTTGACTAGTGTGACAGAAACCTTAAGGCAGTCAGCG 3743
DB	21167 ATTAGGTATGATGACTTGACTAGTGTGACAGAAACCTTAAGGCAGTCAGCG 21226
QY	3744 AACTCGAAAAAGGCGCATACGCTTCAGAAATCGATGAAACACAGCTGCGTAGGAAAA 3803
DB	21227 AACTCGAAAAAGGCGCATACGCTTCAGAAATCGATGAAACACAGCTGCGTAGGAAAA 21286
QY	3804 GAGGCTCAAGAGATATATTGAATAATGAAGATGAACAGATAGTCGCCCAAAAAAGG 3863
DB	21287 GAGGCTCAAGAGATATATTGAATAATGAAGATGAACAGATAGTCGCCCAAAAAAGG 21346
QY	3864 TAAAGAGGCGCAGCAACCAAACTTGTGTGAGGTACCACAAAAGAGAGCAACAAT 3923
DB	21347 TAAAGAGGCGCAGCAACCAAACTTGTGTGAGGTACCACAAAAGAGAGCAACAAT 21406
QY	3924 GAAACCTTCTAAAAAAGGAAGCAAAAAAATGTGACCTCCGACACAGAGAGAGAGGA 3983
DB	21407 GAAACCTTCTAAAAAAGGAAGCAAAAAAATGTGACCTCCGACACAGAGAGAGAGGA 21466
QY	3984 AGAAGAGAAAGCAAAAGTGAATATCGGAATGCAAGATGCAAAAGCAACAGCAGAGT 4043
DB	21467 AGAAGAGAAAGCAAAAGTGAATATCGGAATGCAAGATGCAAAAGCAACAGCAGAGT 21526
QY	4044 GTCAGAGAGACACAGCAGC----- 4062
DB	21527 GTCAGAGAGACACAGCAGCAGTAACTGTAACCTTAACCTGATCTGTTTCGTTAC 21586

QY	4063	-----	-----	AGACGAGA	4070
Db	21587	TATATPATAATCATCAATATTGTGATGCTATCCACATTGGGTCTTCCCAACACAGACGAGA			2164
QY	4071	ATCTCCGGAATCTGTGTCATTTGCATTCACACAGTCCACACACAGAAAGGACGAGAG			4130
Db	21647	ATCTCCGGAATCTGTGTCATTTGCATTCACACAGTCCACACACAGAAAGGACGAGAG			2170
QY	4131	ACCATCAAAAAACGCCATCACCATCATACACAAAAAAATGTG	4173		
Db	21707	ACCATCAAAAAACGCCATCACCATCATACACAAAAAAATGTG	21749		
RESULT 15					
LOCUS	AL138820				
DEFINITION	Human DNA sequence from clone Rpl1-380B4 on chromosome 13, complete				
ACCESSION	AL138820				
VERSION	AL138820.11				
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Emukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 168487)				
TITLE	Smith, M.				
JOURNAL	Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk				
COMMENT	Requests: clonerequest@sanger.ac.uk				
	On Oct 25, 2000 this sequence version replaced gi:10715762.				

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TRERBL; Wp.; WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr13>

RP11-38004 is from the library RP11-11.2 constructed at the Roswell Park Cancer Institute by the group of Pletzer de Jong. For further details see <http://bacpac.med.buffalo.edu/vector/pbac3.6>

This sequence is the entire insert of clone RP11-38004.

```

FEATURES
SOURCE
Location/Qualifiers
1..168487
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-380B4"
/clone_id="RCCI.11.2"
48433 a 33155 c 34624 g 52275 t
BASE COUNT
BRIGIN
```

Query Match	10.7%	Score 447;	DB 9;	Length 168487;
Best Local Similarity	85.5%;	Pred. No. 5,7e-82;		
Matches 550;	Conservative 0;	Mismatches 0;	Indels 93;	Gaps 1;

QY	3624	GTCTGAATTGAGAGAGCCTTGAGGCGAGAAAAAACGCCGTCACAGAACAGSAGSAGAA	3683
Dd	20593	GTCTGAATTGAGAGAGCCTTGAGGCGAGAAAAAACGCCGTCACAGAACAGSAGSAGAA	20652
QY	3684	ATTAGTATGGATGACTTGACTAAGTTGGTCAGGAACAGAAACCTTAAAGGACGTCAGCG	3743
Dd	20653	ATTAGTATGGATGACTTGACTAAGTTGGTCAGGAACAGAAACCTTAAAGGACGTCAGCG	20712
QY	3744	AAGCGGAAAAAGAGGCCATCGGCGCTCAGAACTGTGATGAAACGACGTAGCGCTCAGSAAAA	3803
Dd	20713	AAGCGGAAAAAGAGGCCATCGGCGCTCAGAACTGTGATGAAACGACGTAGCGCTCAGSAAAA	20772
QY	3804	GAGGCTCMAAGAGATATATTACAAAAATGAAGTGAACAGAAATAGTCCGCCCAAAAAAGG	3863
Dd	20773	GAGGCTCMAAGAGATATATTACAAAAATGAAGTGAACAGAAATAGTCCGCCCAAAAAAGG	20832
QY	3864	TAAAAAGAGCCGACACCAAAACCTTGTGTGAGGTACACCAAAAGAAAGACCAACAT	3923
Dd	20833	TAAAAAGAGCCGACACCAAAACCTTGTGTGAGGTACACCAAAAGAAAGACCAACAT	20892
QY	3924	GAAAACTCTTAAAAAAGGAAGCAAAAAAAATGTGGACCTCCAGCACAGSAGSAGAGA	3983
Dd	20893	GAAAACTCTTAAAAAAGGAAGCAAAAAAAATGTGGACCTCCAGCACAGSAGSAGAGA	20952
QY	3984	AGAAAGAAAGAACAAAGTGGAAATACGAAACGAAGTCCAAAGCAAAACGACACCGAGT	4043
Dd	20953	AGAAAGAAAGAACAAAGTGGAAATACGAAACGAAGTCCAAAGCAAAACGACACCGAGT	21012
QY	4044	GTCAGAGAGAGCAGCAGCAG-----	4062
Dd	21013	GTCAGAGAGAGCAGCAGCAGAGTACAGATGTGTAACCTTAACCTGATCTGTTTGCTTAC	21072
QY	4063	-----AGAGCAGA	4070
Dd	21073	TATATTAAATCATATATTGATGCTATCCACACTTGGGTCTTCCCAAGCAGAGCAGA	21132
QY	4071	ATCTCCCGAATCTTAGAGCAATTGAATCCACAGACGTCCACACAGAAAGAGCGSAGAG	4130
Dd	21133	ATCTCCCGAATCTTAGAGCAATTGAATCCACAGACGTCCACACAGAAAGAGCGSAGAGAG	21192
QY	4131	ACCATCAAAAAAGCGCATCACCATCACAGACCAAAAAAAATGTG	4173
Dd	21193	ACCATCAAAAAAGCGCATCACCATCACAGACCAAAAAAAATGTG	21235

Search completed: September 24, 2003, 14:13:31
Job time : 1481 secs

```
Search completed: September 24, 2003, 14:13:31
Job time : 14881 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 24, 2003, 20:22:05 ; Search time 11608 Seconds

(without alignments)
4902.248 Million cell updates/sec

Title: US-09-512-581B-2

Perfect score: 7193

Sequence: 1 MAHSKTRNDCKITPPGVK.....QKGRGRPSKTPSPQPKNV 1391

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-O=/cgn2.1/USPTO_SPOOL/US09512581/runat_23092003_163546_5474/app.query.fasta_1.1543
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=humana0.cdi -LIST=45
-DOCLALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=eto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09512581.ecgn.1_1.7880_etunat_23092003_163546_3474 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_in:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7155	99.5	5309	9	AB023196	AB023196 Homo sapi
2	7155	99.5	7444	9	HS26H231	AL137201 Novel hum
3	7101	98.7	5253	9	HS095825	U95825 Human andro
4	6942	96.5	5317	10	AK102267	AK102267 Mus muscu
5	5659	78.7	5944	10	AK122414	AK122414 Mus muscu
6	4760.5	66.2	6744	9	AF294791	AF294791 Homo sapi
7	2953.5	41.1	1852	9	HS050533	U50533 Human BRCA2
8	2704.5	37.6	5177	9	AB014548	AB014548 Homo sapi
9	2517	35.0	4252	9	BC039256	BC039256 Homo sapi
10	2474	34.4	2265	9	AK098331	AK098331 Homo sapi
11	2474	34.4	2705	9	BC041361	BC041361 Homo sapi
12	2442	33.9	1976	10	BC021408	BC021408 Mus muscu
13	1836	25.5	216608	5	AL844559	AL844559 Zebrafish
14	1729.5	24.0	173613	3	AC007475	AC007475 Drosophill
15	1729.5	24.0	192763	3	AC007474	AC007474 Drosophill
16	1729.5	24.0	292919	3	AB003823	AB003823 Drosophill
17	1708.5	23.8	194634	2	AC020286	AC020286 Drosophill
18	925.5	12.9	2079	9	AK026889	AK026889 Homo sapi
19	925.5	12.9	137246	9	HS49J10	Z84572 Human DNA s
20	925.5	12.9	168487	9	AL138820	AL138820 Homo DNA
21	840.5	11.7	3779	9	BC009650	BC009650 Homo sapi
22	833.5	11.6	4046	8	AB067651	AB067651 Schizosac
23	833.5	11.6	9982	8	SPAC110	AL441624 S.pombe C
24	819.5	11.4	54398	2	AC068224	AC068224 Homo sapi
25	769	10.7	236508	2	AC111126	AC111126 Mus muscu
26	768	10.7	72157	2	AC016449	AC016449 Homo sapi
27	747	10.4	263479	2	AC125986	AC125986 Rattus no
28	706.5	9.8	2212	6	BD159124	BD159124 Primer fo
29	706.5	9.8	2212	6	AK021757	AK021757 Homo sapi
30	698	9.7	439	6	BD059747	BD059747 Secretd
31	697.5	9.7	6411	8	SMA99344	AJ009934 Sordaria
32	645	9.0	772	6	BD145886	BD145886 Primer fo
33	635	8.8	4885	8	EMEBIND	L03200 Emeritella
34	584	8.1	153751	3	AC116551	AC116551 Dictyoste
35	554.5	7.7	54398	2	AC068224	AC068224 Homo sapi
36	523	7.3	184321	2	AC110493	AC110493 Mus muscu
37	523	7.3	190058	2	AC112263	AC112263 Mus muscu
38	523	7.3	298053	2	AC116574	AC116574 Mus muscu
39	520	7.2	233589	2	AC095342	AC095342 Rattus no
40	512	7.1	163355	2	AC117927	AC117927 Mus muscu
41	505	7.0	1541	3	AY058338	AY058338 Drosophill
42	501	7.0	295	6	BD026562	BD026562 Sequence
43	466	6.5	98856	9	AC113150	AC113150 Homo sapi
44	466	6.5	145450	2	AC068352	AC068352 Homo sapi
45	401.5	5.6	85592	2	AL512630	AL512630 Mus muscu

ALIGNMENTS

RESULT 1

AB023196 5309 bp mRNA linear PRI 10-MAY-2002
LOCUS Homo sapiens mRNA for KIAA0979 protein, partial cds.
DEFINITION AB023196
ACCESSION AB023196.2 GI:20521717
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Carniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hiroseawa,M.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
JOURNAL DNA Res. 6 (1), 63-70 (1999)
MEDLINE 99246063
PUBMED 10231032
REFERENCE 2 (bases 1 to 5309)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna@info.kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)
On May 9, 2002 this sequence version replaced gi:4589601.
COMMENT
FEATURES
SOURCE
1..5309
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hj07056s1"
/sex="male"
/tissue_type="brain"
/clone_lib="pluescriptII SK plus"
/dev_stage="adult"
/note="This sequence was obtained by subcloning of the DNA
fragments derived from two cDNA clones (1 - 722 was
derived from fj09522 and 723 - 5309 was derived from
hj07056)."
1..5309
/gene="KIAA0979"
/gene="KIAA0979"
/gene="KIAA0979"
/codon_start=1
/product="KIAA0979 protein"
/protein_id="AA47683.2"
/db_xref="GI:20521718"
/translation="GGGRRRLDAAAADPGEPRSESSSESATRRGRNIVSNASKTRT
NDGKITPPGKVEISDKISKEEMVRRLKMYKTPMDMDQDEEKEKLYLNLALIASD
FELKHPDKVRLAVACLDLIFRIYAEAPTSDDKIDIMFTROLGKHWEDKSD
ENRFYLLLENIAWYKSNICELEDSNEIFOLRTLEYSNNHNGKQVHMMDLMS
SITCEGDTVSOELDPYLVNLYPAHKNLNOQAYLAKALKRTQAIEPTITNFNOY
LMGKTSISDSEHVEPDLIELINIDSHLSLVLPLEFKLSNDNERLDVYLAK
MFGAKDELSAQNPLWQCYLGRFNDHVPRLCEVFAQCLNHDPLADILEYLA
VRSHPDEARHVDIVISVIAAKKIDILNDHNLNFEVERTLDRKWRKXAMGLAQ
IYKKGALOSAGAKDAKOIAIMIKDLIHIYONSIDRLIVERFAQYVMAIFETTE
RMKCLYLYATLIDNAYKALNEMKCONLIRHOVKLIDLILIKOPRTASVAFISKYM
VITRNLPDPRKADQEMKFTVOYLEDEKIRKOLEVLYSPSCSAQACVARETTKAL
MPKOPTNPLEMLIFELERLAPVHIDTESIALIKQVKSIDGTADDEDEVPTDQAI
RAGLELLKVSFTPIISFSAETESILACIKMDEKVAEALQIFKNTGKIEDEP
HIRSLALPVLHHSKSKGPPOAKYAIIHAIHIFSSKEQFOIEPRLKSLDPSNLEH
LITPLVIGHIALAPQAPLAKSLVATFIVKDLNMDRLPGKRTKLAWDEDESP
ETWYKQIATIMMWVILGKNNHSGSTIRLLTTLIHSGLDLEGGKISKSPMSRL
RLAAGSAIVYLAQPCYHEITITLEOYOLCALINDECYONQYAOQLHKSLSRLP
LEHATCALCAKDPKERRARARQCLKNINIVREYIKHOAAVEKSLISLPETVVPY
TTHLADPVPKVDIQLKDYKVECLMKNIVREYIKHOAAVEKSLISLPETVVPY
AOGDDAKMNEKLTVCYDAMNIIIMSKSTYSLESPKDPVLPARFEPDPDNFNTKN
YLPEEMKSTPGPKTNTVYAVNKLPLSSGKOSOTKSSMETEVSNASSSNSSPG
RIGRLDSSEHSENDYMSPLPGKKSPPKRDSDLVRELEKLPGRKKTPTVDEQ
EKIGMDLTITLVQOKPKGSORSKRKRKHTSASESDQEMPEKRLKEDILEDNDQNSP
PKGKRGRPRPLGGTPEKPEPTKTKSGSKSGPAPAEPEEEEEEKOSGNTOKSK

SKORVSRRAQOARESESSAIESSTOSTPOIGRGRPSKTPSPQPKKNVRGRSKOAA
TKENDSSSEYDVFOGSSPYDDIPOEETEEEEEVTVNVRRSKRERRR"
BASE COUNT 1774 a 955 c 1115 g 1465 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5309
Score: 7155.00 Matches: 1387
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 4
Query Match: 99.47% Indels: 0
DB: 9 Gaps: 0
US-09-512-581b-2 (1-1391) x AB023196 (1-5309)
QY 1 MetLAHISerLySThArgThrAsnAspLyLysIleThrTyRProProGlyValLys 20
DB 109 ATGCCTCATTTCAAGACTAGAGCAATGATGAAAAATTACATATCCCGTGGGCTCAAG 168
QY 21 GluIleSerAspLysIleSerLySGluLuketValArgArgLysMetValValLys 40
DB 169 GAATATCATGATTAATAATATCTTAAGAGAGAGATGGTGAACGATTAAGATGTTGTGAAA 228
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyRLeuAsnLeuAla 60
DB 229 ACTTTATGATGATGAGACACGACGACTGCAAGAAAAAGAGCGCTTTATTAAACCTAGCT 288
QY 61 LeuHISLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
DB 289 TTACATCTTGCTTGAGATTTTTTTCACAGCATCTGTAAGATGTTCCCTGCTACTGGTA 348
QY 81 AlaCysCysLeuAlaAspLlePheArgIleTyRAlaProGluAlaProTyRThrSerPro 100
DB 349 GCCTGCTGCTGCTGCTGATATTTTCAGAGATTATCTCTGAGAGCTCCTTACACATCCCT 408
QY 101 AspLysLeuLysAspLlePheMetPheIleThrArgGluLeuLysGlyLeuGluAspThr 120
DB 409 GATTAACATTAAGATATATTTATGTTATTAACAAGACAGCTGAAGGGCTGAGAGATACA 468
QY 121 LysSerProGlnPheAsnArgTyRPhetyrLeuLeuGluLysAsnIleAlaIrpValLysSer 140
DB 469 AAGAGCCACATTCATATAGTATTTTATTTACTTGAGACATCTGCTGGCTGAAGTCA 528
QY 141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyRArgThr 160
DB 529 TATAACATATGCTTGTGATGATGAAGATGCAATGAATTTTACCACAGCTATACGAACC 588
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
DB 589 TTATTTTCAGTTATTAACAATGACACATGCAAGATCCATATGACATGCTGAGACCTT 648
QY 181 MetSerSerIleIleCysGluGluAspThrValSerGlnGluLeuLeuAspThrValLeu 200
DB 649 ATGACCTCTATATTTTGTGAAGAGTGATACAGTGTCTCAGGAGCTTTTGGATACGTTTGA 708
QY 201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyRAspLeuAlaLysAla 220
DB 709 GTAATATCGTAGCTGCTCATTAAGAATTAACAAGACAGATATGATTTGGCAAGAGCT 768
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyRThrIleThrPhePheAsnGlnVal 240
DB 769 TTATGTAAGAGAGACAGCTCAAGCTATATGACGCAATATATTAACATTTTAAAAACAGGT 828
QY 241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 829 CTGATGCTGGGAAAAACATCTCATAGCGATTTTGCAGAGCATGCTTTGACTTAATTTTG 888
QY 261 GluLeuTyRAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 889 GAGCTCAACATATGATGATGATGATGCTGCTGCTGTTTACCCAGCTGAATTTAAA 948
QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300

Db 949 TTAAGAGCAATGATTAATGAGGAGCGCTTACAAGTTGTTAACTACTGCGCAAAATGTTT 1008
QY 301 G1A1A1yAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
Db 1009 GGGGCAAAAGGATTGCAATGGCTTCTCAAAACAAGCCACTTGGCAGAGCTACTGGGCG 1068
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
Db 1069 AGGTTAATGATATCATCTATCCATCCCGCTGGATGTGTCAAAATTTCTGACCATTTCT 1128
QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
Db 1129 CTCATGAACCATCCGATTATACCAAAAGACTTAACAGACTATCTTAAGTGAAGCTACAT 1188
QY 361 ASPProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaLysLys 380
Db 1189 GACCTGAGGAAGCTATTAGACATGTGTATTGTGTCAATGTTACACTGCTGTAAAG 1248
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
Db 1249 GATATTCTTCTGCTCAATGATCATCTTACTTAATTTTGTGAGAGAGAACATTAAGCANA 1308
QY 401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
Db 1309 CGATGGAGAGTACGCAAAAGAACCCATGATGGAGCTGCCCAATTTATGAAGAAATGCT 1368
QY 421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
Db 1369 TTACAGTCAAGAGCTGGAAAGATGTGTCAAAACAGATGATGATCAAGCAAAATGG 1428
QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
Db 1429 CTACATATATATATCAAAATAGTATGTATGTATGCATCTGTTGACAGCATCTTGGT 1488
QY 461 GlnTyrMetValProHisAsnLeuGluThrGluArgMetLysCysLeuTyrTyrLeu 480
Db 1489 CAATCATGTGTTCCCTCACAAATTTAGAAACTACAGAACCGATGAATGCTTATATACTGG 1548
QY 481 TyrAlaTrpLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
Db 1549 TATGCCACAGTGGATTTAAATGCTGTGAAGCATGAAATGAGAAATGTCAAAT 1608
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
Db 1609 CTGCTCCGACATCAAGTAAGAGATTGGCTGACTTGATTAACCAACCCAAAACAGATGCC 1668
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
Db 1669 AGTGTCAAGGCCATATTTTCAAAAGTGATGATTATACAAAGAAATTTACCTGATCCTGCT 1728
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
Db 1729 AAGGCTCAGAGATTATGAAGAAATTCACACAGGTGTAGAGATGATAGAAATAA 1788
QY 561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluLysCysVal 580
Db 1789 AAGCAGTTAAGAGTACTGTGTGTCACACATGCTCTCTGCAACAGCCTCAAGGTGTGTCG 1848
QY 581 ArgGluIleThrLysLysLeuGluLysProLysGlnProThrAsnProPheLeuGluMet 600
Db 1849 CGTGAATATACATAAAGTTGGGCAACCCCAAAACAGCTTACAAATCTTCTGCGAAATG 1908
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
Db 1909 ATCAAGTTTCTCTTGAGAGATACACCTGTGCACATAGATACCAATCTATCAGTGTCT 1968
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
Db 1969 CTTATTAACAAGTAGAACAAATCAATPAGATGACACAGCAGATGAAATGATGAGGTGTT 2028
QY 641 ProThrAspGlnAlaIleArgAlaGluLeuGluLeuLysValLeuSerPheThrHis 660
Db 2029 CCAACTGATCAAGCATCAGACAGGTCTTGAAGTGTAAAGTACTCTCATTTTACACAT 2088

QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Db 2089 CCCATCTCATTTTCATTTCTGTCTGAACATTTGAATCTTACTGGCTTGTGTGAATGAT 2148
QY 681 AspGluLysValAlaGluAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2149 GATGAAGAAAGTAGCAAGAACCTGCACAGCAAAATTTTCAAAAACACAGGAAGCAAAATGGA 2208
QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
Db 2209 GAGGATTTTCCACACATCAGATCAGCTTGTCTTCCGTGTTTACATCACAAATCTAAGAAA 2268
QY 721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
Db 2269 GGACCCCGCTGCMAGCCAAATATGTCATTCATTGTATCATGCCGATATTTCTAGTAAA 2328
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2329 GAGACCCAGTTTGCACAGATATTTGAGCCTGTGATAAAGAGCCTAGATCCAAGCAACCTG 2388
QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2389 GAACATCTCATTAACACCATTTGCTACTATTGGTCATATTGCTCTCTGACCTGATCA 2448
QY 781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2449 TTTGCTGCTCTTGAATCTTTGGTAGCTACTTTTCATTTGTGAAGATTTTCTCATGAT 2508
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
Db 2509 GATCGGCTTCCAGGGGAAAAAGACAACATAACTTTGGGTTCCAGATGAAGAGATCTCCT 2568
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGluMetLys 840
Db 2569 GAGACAAATGTCAAAATTCAGCTATTAAAGATGCTTCGATGGCTACTTGGAAGAAA 2628
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrIleLeuHisSer 860
Db 2629 AATTAATCAGATTAATCAGGAACCTTCTACCTTAAGATTCTTAACAACAATATTGCAAT 2688
QY 861 AspGlyAspLeuThrGluGlnGluLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2689 GATGGAGACTTACAGAACAGGGAAGAAATTAGTAAACAGATATGACAGCTGAGACATT 2748
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleThr 900
Db 2749 GCTGCTGGAGTGTATTTGTGAAGCTGGCACAAGAACCTGTTACCATGAATATCAGCACA 2808
QY 901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
Db 2809 TTAGAACATATATCAGTATGTGCTATGTATATCAACAGATGAATGCTATCAAGTAAACAA 2868
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
Db 2869 GTGTTTGGCCCAAAACTTCCAAAGGCCCTTCCCGTTTACGGCTTCCACTTGATGATATG 2928
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
Db 2929 GCAATCTGTGCTTGTGTCAAAAGATCTGTAAAGAGAGAGAGCTATCTAGGCGAA 2988
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaValSer 980
Db 2989 TGTTTGGTGAAGAAATTAATATATAGGCGGAGATCTTAACAGCATCAGATCTGTAACT 3048
QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3049 GAAAAATATTTGCTCTTACACAGATATGTTGTTCCATATTAACAATTCACCTTTGGCA 3108
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3109 CATGACCCAGATATGTCAAAAGTATACAGATATTAAGTAAACATTAAAGATGTTAAAGATGT 3168

1021 LeuTrpPheValIleuGluIleleuMetaLAllysAsnGluAsnSerHisAlaPheIle 1040
13169 CTTGGTTGTTCTGGAAATATTAATGCTAAAAATGAATAACAGTCAGCTTTTATC 3228
1041 ArgIlySmetValIgluAsnIleIySglInThrIlySAspAlaIglInIlyProAspAspAlaIlys 1060
3229 AGAAGATGTGTGAGAAATATTAAACAAAAGATGCTCCAGAGCCAGATGATCCAAA 3288
1061 MetaSngIlySleuIlyThrValIcySAspValAlaIleIleIleMetSerIlySer 1080
3289 ATGATATGAAAACGTGACACTGTGTGATGCTTGCATGATATCATCATGTCAGAGGT 3348
1081 ThrThrIlySerIleuGluSerProIlySAspProValIleuProAlaIarPhePheThrGln 1100
3349 ACTACATACAGTTGGATCTCTTAAGAACCCGGACTACAGCGCTGTTCTTCTCCTCAA 3408
1101 ProAspIlySAsnPheSerAsnThrIlySAsnIlyIleuProProIgluMetIlySerPhe 1120
3409 CCTGACAGAAATTCAGTAACACCAAAATATCTGCTCCTGAAATGAAATCATTTTTC 3468
1121 ThrProGlyIlySProIlySThrThrAsnValIleuGluAlaValAsnIlySProIleuSer 1140
3469 ACTCTGGAAAACCTTAACACCAATGTTCTAGAGCTCTTAACACCACTTTTCATCA 3528
1141 AlaGlyIlySglInSerGlnThrIlySAspSerArgMetGluThrValSerAsnAlaSerSer 1160
3529 GCAGGCAAGCAATCTCAGACCAAAATCATCAGAAATGAAATGAAACCAAGCAGC 3588
1161 SerSerAsnProSerSerProGlyArgIleIySglIyArgIleuAspSerSerGluMetAsp 1180
3589 AGCTCAAAATCCAAAGCTCTCTGGAAGAAATTAAGGGAGGCTTGAAGTTCTGAAATGGAT 3648
1181 HisSerGluAsnGluAspIlyThrMetSerSerProIleuProGlyIlySAspIlyS 1200
3649 CACAGTGAATTAAGAGATTACACATGCTTCACCTTGCCGGGAAAAAGTGCACAG 3708
1201 ArgAspAspSerAspIleuValArgSerGluIleuGluIlySProArgIlyArgIlySThr 1220
3709 AGAAGCAGCTGTGATCTGTGAAGCTGTGAATTTGAGAGAGCCCTAGAGCGAGAAAAACG 3768
1221 ProValThrIgluIngluIgluIlySleuGluMetAspAspIleuThrIlySleuValIglu 1240
3769 CCCGTCACAGAACGAGAGAAATTAAGTATGCTGACTGACTGACTGACTGACTGACTG 3828
1241 GlnIlySProIlySglInSerGlnArgSerArgIlySArgIlySThrAlaSerGluSerAsp 1260
3829 CAGAACTTAAGGACAGTCAGAGAGTCGAAAGAGCCATACGGCTTCAGAAATCTGAT 3888
1261 GluIgluInThrProIgluIgluIlySArgIleuIySgluAspIleuGluAsnGluAspGlu 1280
3889 GAACAGCAGTGGCTGAGGAGAAAGAGCTCAAAAGAAATATTAAGAAATGAAGATGAA 3948
1281 GlnAsnSerProProIlySglIlySArgIlySArgIlySProProIlySProIleuGly 1300
3949 CAGATAGTCCGCCAAAAAGGTAAGAGAGCCGACACCAAAACCTCTTGCTGGAGGT 4008
1301 ThrProIlySgluIgluProThrMetIlySThrSerIlySglIySerIlySleuSerGly 1320
4009 ACACCAAAAGAGAGCCAAATGAAACTTCTAAAAAGAGAGCAAAAAAAATCTGCA 4068
1321 ProProAlaProGluIgluIgluIgluIgluIgluIgluIgluIgluIgluIgluIglu 1340
4069 CCTCCACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4128
1341 SerIlySerIlySglInIlySArgValSerArgArgIgluIngluIgluIgluIgluIglu 1360
4129 TCCAAAAGCAACAGCACCGAGTGTCAAGGAGACACAGCAGAGAGCGAATCTCTCGAA 4188
1361 SerSerAlaIleGluSerThrGlnSerThrProGluIlySglIyArgIlySArgIlyS 1380
4189 TCTAGTGCATTTGATTCACACAGTCCACACACAGAAAGAGCAGGAGGAGCATCAAAA 4248
1381 ThrProSerProSerGlnProIlySlysAsnVal 1391

DB 4249 ACCGATCACCATCAACCAAAAAAATGTC 4281
RESULT 2
LOCUS HS26H231 7444 bp mRNA linear PRI 25-JAN-2000
DEFINITION Novel human gene mapping to chromosome 13.
ACCESSION AL137201
VERSION AL137201.1 GI:6759511
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Rhodes, S. and Huckle, E.
TITLE Submitted (13-JAN-2000) E-mail contact: humquerry@sanger.ac.uk
JOURNAL This cDNA sequence was assembled from public domain ESTs and single pass sequencing reads from expressed DNA templates, aligned to the genomic DNA sequence from the bacterial clones 26H23 (284467), 267P19 (275889) and 49J10 (284572).
COMMENT The EST sequences listed match this sequence with an identity of at least 95% between the coordinates shown.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene Sanger Centre name : 26H23.C13.1.
FEATURES
source
location/Qualifiers
1. 7444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12.3"
1. 107
/number=1
108. 234
/number=2
127. 4470
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB69911.1"
/db_xref="GI:6759512"
/translation="MAHSKRTNDGKITTPPGYKEISDKISKEMVRRLKMYTFMD
MDQSEEEKELYLNLALHLASDFELKHPDVRILVACCLADIRIAPAPAYSPDK
LKDFMFITRLKLEDPKSPQENRYFLLENLWVSYNCFELDESNELFPQLYPT
LFESYINNGKVMHMYHMDLMSSTICGDPYSEDLDTPTVLVNIYPAKRNKQAYDA
KALKRTQAQIEPTITNPFNVYMLKRTSISDSHFYDLLELYNIDSHTLSVLPQ
LEFLKSNDSNEERLQVVKLAKMFGAROSELASQKPLMOCYLGKRFNDHVPRLKEV
KFASHCLMNHEDLAKDLLEYLAKVSHDPEERIRHDVIVSYTAAKDLILVNDHLNF
VRERTTLKRMRVRKEAMMGLAOIKRYALASAAKDAKQIAIKDKLILHYONSID
DLILVERIFYQVYVPHNLETERMKCYLYATATDLNAVKALNEMKCOMLIRGOVD
LIDLIRKOPDASVKALFESKVMVTRNLPDGKQDMKRTQVLEDEKIRKOLEVL
VSPKSCQKQARCYRETKRKNKQPNPRLBKIKLELIAVNHIDTSISALIQ
VKSIDIGTADDEDGVPIDQAIRAGLEULKLSHFTISFSAETFEISLALWDESS
KVAEALQIEFNKTSKLEEDPHIRSLALVYLHKSKGPPRQAKVAIHCIALWSSK
ETQPAQIETPEPLHKSLDPSNLEHLITPTVLTGILALPDQFAPLKSLVFTYKDL
MNDRLPGKTKTKLWNPDEEVSPEPTWVIOAIKMWVNLGKKNHSGSTSLKLT
ILHSDGILEGKTSKPDMSRLRLAAGSALVYKLAOEPICYHEHITLBOYCALAINE
CYOYROYFAOKLHKGLSRLRPLEYMAICALCAKDPKERRAARQCLVKNIRREY
LKQHAAYSEKLSLPLEYVYTHLADHDVYKVDLIDLKVKRCLEWVLEITLMA
KNENNSHAFIRKVENIKOTKDAQPDADKNEKLYVCDVAMNIIMSKSTYSLESP
KDPYLPAFEFQDPKNSNTKNYLPPEMKSPFTGKTKTIVLGAVKPLSAGKQSQ
TKSRRMEVSNASSSSSPGRILKGLSDSEMDHSSNEDYTMSPLPGKSKDRDS
DLVSELEKPRGKRKTPYTBDEKLGMDLITKLYODKPKGSQSRKRGKSHASDQ
OWPEKRLKEDILENEDQNSPPKKGRRGPPKPLGGGTPEEPTMKTSKGSKKGG
PPAPEEEEEEERQSGNTEOKSKSOHVSRAAOQRASSPESALIESQTPQKGRGP
SKTSPSPQPKKNVYVGRSKQAATKENDSSEEDVDFGSSPVDIDIPQETEESEVSTVN
VRRSAKRERR"

```

exon          526..623
              /number=5
exon          624..750
              /number=6
exon          751..831
              /number=7
exon          832..972
              /number=8
exon          973..1088
              /number=9
exon          1089..1183
              /number=10
exon          1184..1329
              /number=11
exon          1330..1481
              /number=12
exon          1482..1595
              /number=13
exon          1596..1677
              /number=14
exon          1678..1726
              /number=15
exon          1727..1866
              /number=16
misc_feature  1866..2046
              /note="matches EST AA687846 from clone IMAGE:1219676"
misc_feature  1866..1941
              /note="matches EST AA687776 from clone IMAGE:1219628"
exon          1867..1982
              /number=17
exon          1983..2088
              /number=18
exon          2089..2249
              /number=19
exon          2250..2373
              /number=20
exon          2374..2532
              /number=21
exon          2533..2601
              /number=22
exon          2602..2738
              /number=23
exon          2739..2862
              /number=24
exon          2863..3067
              /number=25
misc_feature  3067..2912
              /note="complement(2912..3302)"
exon          3068..3182
              /note="matches EST AA630616 from clone 856164"
exon          3183..3315
              /number=26
exon          3302..3572
              /note="matches EST AA333118"
misc_feature  3316..3435
              /number=28
exon          3436..3498
              /note="matches EST AA282190 from clone IMAGE:704631"
misc_feature  3499..3644
              /number=29
exon          3645..3750
              /number=30
misc_feature  3751..4190
              /note="complement(3586..4065)"
exon          4001..4119
              /note="matches EST A1655429 from clone IMAGE:2239331"
misc_feature  4135..4450
              /number=32
              /note="complement(3864..4286)"
              /note="matches EST A1680124 from clone IMAGE:2264521"
              /note="matches EST N83765 from clone KK2745"

```

```

exon          /note="matches EST F12354 from clone C-38h11."
              4191..4298
              /number=33
misc_feature  4255..4307
              /note="matches EST T64867 from clone 22264"
exon          4299..4434
              /number=34
misc_feature  4300..4518
              /note="matches EST H44781 from clone 188032"
exon          4435..7444
              /number=35
misc_feature  4537..5022
              /note="matches EST A1681422 from clone IMAGE:2272632"
misc_feature  4537..4892
              /note="matches EST AA911125 from clone IMAGE:1416862"
misc_feature  4538..4969
              /note="matches EST A1275866 from clone IMAGE:1877482"
misc_feature  4538..4947
              /note="matches EST AA987361 from clone IMAGE:1603350"
misc_feature  4538..4943
              /note="matches EST A1299646 from clone IMAGE:1898010"
misc_feature  4562..5003
              /note="matches EST H93424 from clone 220700"
misc_feature  4571..4854
              /note="matches EST W20070 from clone 306049"
misc_feature  4746..4994
              /note="matches EST AW089775 from clone IMAGE:2593146"
misc_feature  4746..4994
              /note="matches EST AA747568 from clone IMAGE:1269696"
misc_feature  4803..5332
              /note="matches EST A1911784 from clone IMAGE:2329873"
misc_feature  4828..5296
              /note="matches EST AA563884 from clone IMAGE:1013990"
misc_feature  4861..5243
              /note="matches EST D20082 from clone pml283."
misc_feature  4872..4921
              /note="matches EST AA17347 from clone 621810"
misc_feature  4872..4921
              /note="matches EST AW212954 from clone IMAGE:2647801"
misc_feature  4872..4921
              /note="matches EST AA154932 from clone 599194"
misc_feature  4872..4921
              /note="complement(4872..4921)"
misc_feature  4872..4921
              /note="matches EST AW048929 from clone
              UT-M-BH1-amo-C-09-0-UT"
misc_feature  4876..5333
              /note="matches EST A1016896 from clone IMAGE:1627886"
misc_feature  4876..5333
              /note="complement(join(4922..4975,5281..5332))"
misc_feature  4876..5333
              /note="matches EST A1012649 from clone RPLBNO2"
misc_feature  4876..5333
              /note="complement(4922..4975)"
misc_feature  4876..5333
              /note="matches EST AA899586 from clone
              UT-R-E0-cy-e-02-0-UT"
misc_feature  4876..5333
              /note="matches EST A1172172 from clone RMDBV82"
misc_feature  4876..5333
              /note="complement(4931..5332)"
misc_feature  4876..5333
              /note="matches EST AA625961 from clone 745478"
misc_feature  4876..5333
              /note="complement(4951..5332)"
misc_feature  4876..5333
              /note="matches EST A1140790 from clone IMAGE:1566222"
              /note="complement(5021..5309)"

```

Alignment Scores:

```

Score: 0 7444
Percent Similarity: 99.718
Best Local Similarity: 99.718
Query Match: 99.478
DB: 9 Gaps: 0

```

US-09-512-581B-2 (1-1391) x HS26H231 (1-7444)

```

Oy      1 MetAlaHhSserLySthrArGrThrAsnAspGlyIlySleHrTyrProProGlyValLys 20
        |||
        127 ATGGCTCATTTCAAGACATGACCAATGATGAAAAAATTACATATCCGCTGGGGTCAAG 186

```

QY	21	GIUIESerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys	40
Db	187	GAATATATCATTAATAATATCTAAAGAGAGATGGAGACGATTAAGAATGCTGTGA	246
QY	41	ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla	60
Db	247	ACTTTAAGATATGACGACGAGACTCTGAAGAAGAAAGAGCTTTATTTAAACCTAGCT	306
QY	61	LeuHISLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal	80
Db	307	TTACATCTGCTGCTGAGATTTTTTTTCTCAGACATCTGTATTAAGATGTTCCGTTACTG	366
QY	81	AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro	100
Db	367	GCCGCTGCTCCCTGCGATATTTTTCAGAGATTTATGCTCTCGAAGCTCTTACACATCCCT	426
QY	101	AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr	120
Db	427	GATTAACATTAAGGATATTTATGTATTAAACAAGACGTTGAAGGGGCTTAAGAGATACA	486
QY	121	LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTyrValLysSer	140
Db	487	AAGAGCCCAATTCATTAAGTATTTTTTATTACTTGAGAACATTCCTGGGCTCAAGCA	546
QY	141	TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr	160
Db	547	TATTAACAATGCTCTTGAAGTATGAAGATGCAATGAATTTTCACCCAGCTTATACAGAAC	606
QY	161	LeuPheSerValIleAsnAsnGluHisAsnGlnLysValHisMetHisMetValAspLeu	180
Db	607	TTATTTTACGTTATTAACAATGGCCACATATAGAAAGTCCATATGCACATGTAGACCTT	666
QY	181	MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu	200
Db	667	ATGAGCTCTATTTATTTGAGAGGTGATACAGTCTCTCAGAGCTTTTGATACGCTTTTA	726
QY	201	ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla	220
Db	727	GTTAAATCGGATCCGCTCATAGACATTTTAAACAAGCAGCATATGATTTGGCAAGGCT	786
QY	221	LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleProThrPhePheAsnGlnVal	240
Db	787	TTACTGAGAGAGACAGCTCAGCGATTTAGCGCATATATTACCATTTTTTTTAATCAGGTT	846
QY	241	LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu	260
Db	847	CTGATGCTTGGGAAAAACATCTATACAGCATTTGTCAGAGCATTGCTTGAATTAATTTTG	906
QY	261	GluLeuTyrAsnIleAspSerHisLeuLeuLeuLeuSerValLeuProGlnLeuGluPheLys	280
Db	907	GAGCTCTACATATATGATAGTCACTTTGGTCTCTCTGTGTTTAAACCCAGCTTGAATTTAAA	966
QY	281	LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe	300
Db	967	TTTAAAGACATGATTAATAGAGAGCCGCTACAACTTGTTAACTACACGCAAAATGTTT	1026
QY	301	GlyAlaLysAspSerGluLeuAlaISerGlnAsnLysProLeuTyrPginCysTyrLeuGly	320
Db	1027	GGGGCAAAAGATTCAGATTTGGCTTCTCAAAACAAGCCACTTTGGAGTGTACTTGGGC	1086
QY	321	ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys	340
Db	1087	AGGATTAATGATATCCATGTACCAATCCGCTGGAAATGTGAAATTTGCTAGCCATTTGT	1146
QY	341	LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis	360
Db	1147	CTCATTAACAACATCTCTGATTTAGCAAAAGACTTAAACAAGATATCTTAAAGTAGGTCACAT	1206
QY	361	AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys	380
Db	1207	GACCCGTGAGAGATATTAGACATCAAGATGTTATTTGTCTCAATAGTTACAGCTGTCAAAAG	1266

QY	361	ASPILELeuLeuValAsnAspHisLeuLeuAsnPhetValArgGluArgThLeuAspLys	400
Db	1267	GATATTTCTTCGGTCGAATGATCACACTTACTTAATTTTGGAGAGAGAACTTTGACAAA	1326
QY	401	ArgTPArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyLysLysTyTala	420
Db	1327	CGATGGAAAGTACGCAAGAAAGCCATTGATGGACTTGCCTCAAAATTTATAGAAATATGCT	1386
QY	421	LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTPIleLysAspLysLeu	440
Db	1387	TTACAGTCGACGAGCTGGAAGAAGATGCTCCAAACAGATAGCATGGATCAAGACAAATTG	1446
QY	441	LeuHisIleTyTYTYGlnAsnSerIleAspSparGleuLeuValGluArgIlePhetAla	460
Db	1447	CTACATATATATTTATCAAAATAGATATCATGATGACACTTGTGTGAAGGATCTTGTCT	1506
QY	461	GlnTYMetValProHisAsnLeuGlnThrGluArgMetLysCysLeuTYTYTLeu	480
Db	1507	CAATTCATGTTCTCTCACAAATTTAGAACTACAGACGAGTGAATGCTTATATTAATTG	1566
QY	481	TyAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGlnMetTyrLysCysGlnAsn	500
Db	1567	TATGCGACACATGAGATTAAATAGCTGTGAACACATTGAATGAATGTGAATGTCAAAAT	1626
QY	501	LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThAspAla	520
Db	1627	CTGCTCCACATCAAGTAAAGGATTTGCTTACATTGATTAAGCAACCCAAATACAGATCC	1686
QY	521	SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
Db	1687	AGTGTCAAGGCCATATTTTCAAAAGTGATGTTATTACAGAAATTTTACCTGATTCGTGT	1746
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnAspAspGluLysIleArg	560
Db	1747	AAGGCTCAGGATTTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAATAGA	1806
QY	561	LysGlnLeuGluValLeuValSerProThrySerCysLysGlnAlaGluGlyCysVal	580
Db	1807	AAGCAGTTAGAGACTCTTGTTAGTCCAACTGCTCCGCAAGCGGCTGAAGGTTGTCTGT	1866
QY	581	ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGlnMet	600
Db	1867	CGTGAATTAACCTAGAAAGTTGGGCAACCCCAACAGCCTACAAATTCCTTCTGTGAATG	1926
QY	601	IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla	620
Db	1927	ATCAAGTTCTTCCTGGAGAGATFACACCTGTGCACATAGATACCAATCATCATAGTCT	1986
QY	621	LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal	640
Db	1987	CTTATTTAAACAAGTGAACAAATCATATGATGCAACAGCATGTGTAAGATGAGAGGTCTT	2046
QY	641	ProThrAspGlnAlaIleArgAlaGlyLeuGlnLeuLeuLysValLeuSerPheThrHis	660
Db	2047	CCAACTGATCAAGCCATCAGACAGAGCTTAAACGCTTAAAGTACTCTCATTTTACAT	2106
QY	661	ProIleSerPheHisSerAlaGlnThrPheGluSerLeuLeuAlaCysIleLysMetAsp	680
Db	2107	CCCATCTCATTTCTCATTTCTGTGGAACATTTTAATCATTAATTCAGCGCTGTCTAAAAATGAT	2166
QY	681	AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu	700
Db	2167	GATGAAAAGTAGCGAAGAGTGCACCTACAAATTTTCAAAAACACAGGAAGCAAAATTTGAA	2226
QY	701	GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys	720
Db	2227	GAGGATTTTCCACACATCAGATCAGCCCTTCTCTGTATTACATCAACAAATTCAAAAA	2286
QY	721	GlyProProAspGlnAlaLysTyTalaIleHisCysIleHisAlaIlePheSerSerLys	740
Db	2287	GGACCCCCCGTCAAGGCCAAATATAGCCATTCATTGTATCATGAGGATATTTTTCAGTAA	2346
QY	741	GluThrGlnPheAlaGlnIlePheGlnProLeuHisLysSerLeuAspProSerAsnLeu	760

|||||
Db 2347 GAGACCAAGTTGGACACAGATATTGAGCCTGTGACATAAGAGCCATGATCCAAAGCAACCTG 2406
Qy 761 GUNHSLSEUILLethProleuValThrlleGlyHISleAlaleuLeuAlaProaspGln 780
Db 2407 GAACTATTCATTAACACCATGGTGTATGGTCAATATTGCTCTCTCCACCTGTATCA 2466
Qy 781 PheAlaAlaProTrpLysSerTrpAlaIaThrPheIleValLysAspLeuMetAsn 800
Db 2467 TTTGCTGCTCCCTTGGAAATCTTTGGTAGCTACTTTTCATTGTGAAGATCTTTCATGAT 2526
Qy 801 AsparGleuProGlyLysLysThrThrlLysLeuTrpValProaspGluValSerPro 820
Db 2527 GATCGGCTCCAGGAGAAAAAGACAACATACTTGGGTTCCAGATGAAGAACATCTCT 2586
Qy 821 GluTrpMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuLys 840
Db 2587 GAGAACATGCTCAAAATTCAGGCTATTAAATGATGGTTCGATGGCTACTTGGAATGAAA 2646
Qy 841 AsnAsnHISerLysSerGlyThrSerThrlLeuArgLeuLeuThrlLleuHISser 860
Db 2647 AATATTCACAGTAATTCAGAACTTCTACCTTAAGATTGCTTAACAACATATTGCAATG 2706
Qy 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProaspMetSerArgLeuArgLeu 880
Db 2707 GATGAGAGACTTGACAGAACAGGGAATAATTAGTAACACGATATGTCACGCTGAGACT 2766
Qy 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProGlyThrlHISgluIleThr 900
Db 2767 GCTGCTGGAGCTGCTATTGTGTGAAGCTGCACAAAGCCCTGTACCATGAAATCATCA 2826
Qy 901 LeuGluGlnTrpGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrlGlnValArgGln 920
Db 2827 TTAGAACAATATTCACACTATGTGCATTAGCTATCAACGANTGATGCTATCAATAAGACA 2886
Qy 921 ValPheAlaGlnLysLeuHISLysGlyLeuSerArgLeuArgLeuProLeuGluTrpMet 940
Db 2887 GTGTTGGCCAGAAACCTTCACAAAGCCCTTCCGTTTACGCGCTTCCACTTGAGTATATG 2946
Qy 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHISAlaArgGln 960
Db 2947 GCATCTGTGCCCTTTGTGCAAAAGATCCCTGAAGAGAGAGAGAGCCATCTATGCGCA 3006
Qy 961 CysLeuValLysAsnIleAsnValArgArgGluTrpLysGlnHISAlaAlaValSer 980
Db 3007 TGTGTGGTGAATAATATATGTAGGGGGGATCTGAAGCACACATGSCACTTTTACT 3066
Qy 981 GluLysLeuLeuSerLeuLeuProGluTrpValValProTrpThrlLleHISleuLeuAla 1000
Db 3067 GAAAAATTAATGCTCTCTACACAGATATGTTGCCATATACAAATTCACCTTTTGGCA 3126
Qy 1001 HISAspProaspTrpValLysValGlnAspIleGlnGluLysAspValLysGluCys 1020
Db 3127 CATGACCCAGATATATGTCAAGATACAGATATTGAACAACCTTAAGAGTTTAAAGATGT 3186
Qy 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHISAlaPheIle 1040
Db 3187 CTTGGTTTGTCTGTGAATAATTATATGGCTAAATAAGAAATACAGTCACGCTTTTATC 3246
Qy 1041 ArgLysMetValGluAsnIleLysGlnThrlLysAspAlaGlnGlyProaspAspAlaLys 1060
Db 3247 AGAAGAGTGTGAAGAAATATTAAACAAACAAAGATGCCCAAGGACAGATATATCAAAA 3306
Qy 1061 MetAsnGluLysLeuTrpThrlValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3307 ATGATATGAAAAACTTACACTGTGTGATGTGTCATGAATATATCATCATGCAAGACT 3366
Qy 1081 ThrThrlTySerLeuGluSerProLysAspProValLeuProAlaArgPheThrlGln 1100
Db 3367 ACTACATACAGATTGGAAATCTCTTAAGACCCGAGTACTACGAGCTCGTTTCTTCACTCA 3426
Qy 1101 ProaspLysAsnPheSerAsnThrlLysAsnTrpLeuProProGluMetLysSerPhe 1120
|||||

Db 3427 CCTGACAGAAATTTAGTAACACCAAAATTAATCTGCTCTGAAATGAATCATTTTTC 3486
Qy 1121 ThrProGlyLysProLysThrlTrpAsnValLleuGlyAlaValAsnLysProLeuSer 1140
Db 3487 ACTCTGGAAAACTTAAACCAACCAATGTTCTAAGGCTGTTAACAGCACCTTTTCATCA 3546
Qy 1141 AlaGlyLysGlnSerGlnThrlLysSerSerArgMetGluThrlValSerAsnAlaSerSer 1160
Db 3547 GCAGCCAGCAATCTCAGACCAATCATACCAATGGAACACTGAAGCAATGCACAGCAC 3606
Qy 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3607 ACCTCAAAATCCAAAGCTCTCTCGAAGAAATAAAGGAGGAGCTTGATATGTTCTGAAATGAT 3666
Qy 1181 HISSerGluAsnGluAspTrpThrlMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3667 CACAGTGAATAAGAAATTAACCAATGCTTCACTTGGCCGGGAAAAAAGTGCAAG 3726
Qy 1201 ArgAspAspSerAspLeuValArgSerGluLeuGlyLysProArgGlyArgLysLysThr 1220
Db 3727 ACAGACGACTGATCTTGTAAAGTCTGAATTTGAGAACCCCTAGAGCCGAGAAAAAAG 3786
Qy 1221 ProValThrlGluGlnGluLysLeuGlyMetAspAspLeuThrlLysLeuValGlnGlu 1240
Db 3787 CCCGTCACAGAACAGAGAGAAATTAAGTATGATGACTTGACTTAAGTTGGTACAGCA 3846
Qy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHISThrlAlaSerLysSerAsp 1260
Db 3847 CAGAAACCTTAAAGGAGTCAAGCAAGTCCGAAAAAGGCCATACGCTTCAAAATCTGAT 3906
Qy 1261 GluGlnGlnTrpProGluLysLysArgLeuLysGluAspIleLeuGluLysAsnGluAspGlu 1280
Db 3907 GAAACGAGCTGCTGAGAGAAAAAGAGCTCAAAAGAAATATTGAATAATCAAGATGA 3966
Qy 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyLysGly 1300
Db 3967 CAGAAATAGCCGCCCAAAAGGTTAAAGAGGCCACCAACCAAACTTGTGTGAGAGT 4026
Qy 1301 ThrProLysGluGlnProThrlMetLysThrlSerLysGlySerLysLysSerGly 1320
Db 4027 ACACCAAAAGAAAGACCAACATGAATACTTTAAAAAAGAAAGCAAAAAAAATCTGCA 4086
Qy 1321 ProProAlaProGluGluGlnGluGluGluArgGlnSerGlyAsnThrlGluGlnLys 1340
Db 4087 CCTCAGCAGCAGAGAGAGAGAGAGAAAGCAAAAGTGAATATCGGAACAGAAAG 4146
Qy 1341 SerLysSerLysGlnHISArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4147 TCCAAAAGCCAAACAGCACCGAGTGTCAAGGAGACACAGACAGCAAGCAAAATCTCTGAA 4206
Qy 1361 SerSerAlaIleGluSerThrlGlnSerThrlProGlnLysGlyArgGlyArgProSerLys 1380
Db 4207 TCTAGTGCATTAATTCACACAGATGCCACACAGAAAGGACGAGAAAGCAATCAAAA 4266
Qy 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4267 ACGGCATCACCATCAACACCAAAAAAAATATGTC 4299
RESULT 3
HSU95825 5253 bp mRNA linear PRI 05-APR-1999
LOCUS Human androgen-induced prostate proliferative shutoff associated
DEFINITION protein (AS3) mRNA, complete cds.
ACCESSION U95825
VERSION U95825.2 GI:4559409
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
REFERENCE
1 (bases 1 to 5253)
Geck, P., Szelei, J., Jimenez, J., Soto, A.M. and Sonnenschein, C.
Androgen-induced proliferative shutoff in prostate cancer cells

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
SOURCE
Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
2 (bases 1 to 5253)
Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
Early gene expression during androgen-induced inhibition of
proliferation of prostate cancer cells: a new suppressor candidate
on chromosome 13, in the BRCA2-R1 locus
J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
99229875
10215036
3 (bases 1 to 5253)
Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
Direct Submission
Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
4 (bases 1 to 5253)
Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. and Soto, A.M.
Direct Submission
Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
Sequence update by submitter
On Apr 5, 1999 this sequence version replaced gi:4539617.
location/Qualifiers
1..5253
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
/cell_line="LNCap"
/cell_type="cancer cells"
/tissue_type="prostate"
/note="exon 1 maps to p1 Artificial Chromosome PAC26H23;
exons 2-24 map to cosmid 267p19; exons 25-34 map to p1
Artificial Chromosome PAC 49j10"
1..5253
/gene="AS3"
1..46
/gene="AS3"
/number=1
47..173
/gene="AS3"
/number=2
66..4241
/gene="AS3"
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein"
/protein_id="AAD22134.2"
/db_xref="GI:4539410"
/translation="MAHSKTRNDGKITPPGVKELSDKISKEMVRRLKMYVTEND
MODDSEEEKLYNLALHLASDFELKHPDKDVLVACCLADIFRIYAPAPYTSPPK
LKDIPEITRQLGLEEDTKSPFNRYFLLENTAWSYNICPELEDSNLTFTQLYST
LFSVINGNHQKVMHVMVLDSSIIICGVOSELDTVYVNLVPARKINLQAYDA
KALKRQAQALEPYITTFNOVLGKTSISDSEHVSIDLILELYNDSHLVLPQ
LEFKLSNDNEERLQVVKLLAKMGAGDSLASQNKPLWCQYLGREFNDIHPRIELCY
KFSASHCMNHDLAKDTEYLKVSHPBEAIRDVIVSVTAAKDILVNHDLNF
VGERTLDKRVRREKAMGLAQIYKYLASAGKDAKOIATIKDKLHLYONSID
DRLVERIFAYVYVPHNLFTETTERKCLYIYATLIDNAVKALNEMKCOMLROVND
LIDLKOPKTDASVKAFESKVMYITRSLPPGKAQODPMKFTQVLEDEKTROLEV
VSPTCCKQAEVCEVREITKRLGNKOPTNPFLEKTLERLAVHIDTSSISLKO
VNSIDGTADDEDEGVPTDQIRAGLELKLVSFTHPSHSAETFESESLACLMDSE
KVAEALQIFKNTGSKIEEDPHIRSALEPLVHLHKSCKGPPROAKYAHCIIHAFSSK
EAOFAQIEPLKSLDPSNLEHLITPLVTIGHTALLADPOFAPLKSLVNFETKDL
MMDRLPGKTTKTLMPDEEYSPETMYIOAIKMMVRLGLMKNHKSQSTLRILT
ILUSDGDLTQGGKISKPDMSRLAAGSATYKLAQECYHEIITLBOYQICALAINDE
CIVQOVFAQLKHLGSLRLPLEYMAICALCAKDPYKERRAHARQCLVKNINVRREY
LKQHAAYSEKLLSLLEPYVVPYIHLAHPDVVYKODIQLDKVRECLVFEILWA
KNENSHAFIRKVENIKQTKDAQGPDADKNEKLYVCDVAMNIIKSKSTYSLESP
KDVLPAEFQTDKNESNTKNYLPCKMSFTPGKPKTTNVLGAVNKPISAGKOSQ
TKSSMETVGNASSSSNPSRGRIGKGLDSSMDHSENYTMSSPILPGRKSRDQ
DIVRSELEKPRGKKTPTVREKELGMDDTKLVOBRPGRGSRKRGHTASSSDQ
QMPBEKRLKEDIENEDONSPPKGRGKRPPLVGGGTPEEPTMTKTSKGSKKKG
PPAPEEEEREQSGNTEQSKSKQHVRRAOQRAESSSSAIESSTQSTPKQGRGP

SKTSPSPQPKKNV"
174..377
/gene="AS3"
/number=3
378..464
/gene="AS3"
/number=4
465..562
/gene="AS3"
/number=5
563..689
/gene="AS3"
/number=6
690..770
/gene="AS3"
/number=7
771..911
/gene="AS3"
/number=8
912..1027
/gene="AS3"
/number=9
1028..1122
/gene="AS3"
/number=10
1123..1268
/gene="AS3"
/number=11
1269..1420
/gene="AS3"
/number=12
1421..1534
/gene="AS3"
/number=13
1535..1616
/gene="AS3"
/number=14
1617..1665
/gene="AS3"
/number=15
1666..1805
/gene="AS3"
/number=16
1806..1921
/gene="AS3"
/number=17
1922..2027
/gene="AS3"
/number=18
2028..2188
/gene="AS3"
/number=19
2189..2312
/gene="AS3"
/number=20
2313..2471
/gene="AS3"
/number=21
2472..2540
/gene="AS3"
/number=22
2541..2677
/gene="AS3"
/number=23
2678..2801
/gene="AS3"
/number=24
2802..3006
/gene="AS3"
/number=25
3007..3121
/gene="AS3"
/number=26

exon 3122.. 3254
/gene="AS3"
/number=27
exon 3255.. 3374
/gene="AS3"
/number=28
exon 3375.. 3437
/gene="AS3"
/number=29
exon 3438.. 3583
/gene="AS3"
/number=30
exon 3584.. 3689
/gene="AS3"
/number=31
exon 3690.. 4129
/gene="AS3"
/number=32
exon 4130.. 4354
/gene="AS3"
/number=33
exon 4355.. 4810
/gene="AS3"
/number=34

BASE COUNT 1755 a 944 c 1074 g 1480 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 5253
Score: 7101.00 Matches: 1377
Percent Similarity: 99.28% Conservative: 4
Best Local Similarity: 98.99% Mismatches: 10
Query Match: 98.72% Indels: 0
DB: 9 Gaps: 0

US-09-512-581B-2 (1-1391) x HSU95825 (1-5253)

QY 1 MetLAHISserLysThrArgThrAsnAspGlyLysIleThrTyrProGlyValLys 20
DB 66 ATGGCTCATCAAGACTAGAGCCATATGTAAGAAATTTACATATCCCGCTGGGGTCAAG 125
QY 21 GluIleSerAspLysIleSerLysGluIleValArgArgLeuLysMetValLys 40
DB 126 GAATATACAGATTAATAATCTAAAGAGAGATGGTGAACGATTAAAGATGGTGTGAAA 185
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeuAla 60
DB 186 ACTTTATGATATGACACGAGACTCTGAAGAGAAAGAGAGCTTTATTAAACCTAGCT 245
QY 61 LeuHISleuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuVal 80
DB 246 TTACATCTTGCTTCAGATTTTCTCAAGCATCTGTAAAGATGTCGCTTACTGTA 305
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
DB 306 GCCGTGCTGCTTGGCGATATTTTCAGATTTATGCTCTGAAGCTCCCTTACACATCCCT 365
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGluLeuGlnAspThr 120
DB 366 GATAAACTAAAGATATATTTATCTTTAAACAAACAGTGAAGGGCTGAGAGATACA 425
QY 121 LysSerProGlnIlePheAsnArgTyrPheTyrLeuLeuGlnAlaIleAlaTyrValLysSer 140
DB 426 AAGAGCCACAGTTCAATAGGATTTTATTTGCTTGAGAACATTGGCTTGCAAGTCA 485
QY 141 TyrAsnIleCysPheGluLeuGlnAspSerAsnGlnIlePheThrGlnLeuTyrArgThr 160
DB 486 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACGAACC 545
QY 161 LeuPheSerValIleAsnAsnGlnLysAsnGlnLysValHisMetHisMetValAspLeu 180
DB 546 TTATTTTCAGTTATTAACAATAGCCACATACGAAAGTCATATGACATGTAAGACCTT 605

QY 181 MetSerSerIleIleCysGluGluYAspThrValSerGlnGluLeuLeuAspThrValLeu 200
DB 606 ATGAGCTTATTTATTTTGGAAGGTGATACAGTGTCTCGAGAGCTTTTGGATACGGTTTAA 665
QY 201 ValAsnLeuValProIleHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
DB 666 GTAAATCTGGTACCTGCTCATAGAAATTTAAACAAGCAAGCATATGATTTGGCAAGGCT 725
QY 221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
DB 726 TTACGGAAGAGACAGCTCAAGCTAATTGAGCCATATATTACCAATTTTAAATCAGGTT 785
QY 241 LeuMetLeuGluLysThrSerIleSerAspLeuSerGlnHisValPheAspLeuIleLeu 260
DB 786 CTGATGCTTGGGAAAACCTCTATCAGCGATTTGTCAGAGCATGTCTCGACTTAATTTTG 845
QY 261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
DB 846 GAGCTTCACATATATGATAGTCATTTGCTGCTCTCTGTGTTTACCCAGCTTGAATTTAA 905
QY 281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
DB 906 TTAAAGAGCAATGATTAATGAGGAGCGCTTACAACTTGTAAACTAGTGGCAAAATGTTT 965
QY 301 GluAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTyrProGlnCysTyrLeuGly 320
DB 966 GGGCGCAAGAGATTTCAGATTTGGCTTCTCAAAACAAGCCACTTTGGCACTGCTACTGGGC 1025
QY 321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
DB 1026 AGGTTTATGATATTCATCATGACATCCGCTGGAATGTGGAATTTGCTAGGCATGT 1085
QY 341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
DB 1086 CTCATGAAACCATCTCGATTTTGAAGAAACCTTAAACAGATGTCTTAAAGTGAAGTGCAT 1145
QY 361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
DB 1146 GACCTTGAGGAAGCTATTAGACATGATGTTTGTGTCAATATAGTTACAGCTGCTAAAAAG 1205
QY 381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
DB 1206 GATATTTCTTGGTCATATGATCATTACTTAAATTTTGGGAGAGAAACATTTAGACAAA 1265
QY 401 ArgTyrPheValArgLysGluAlaMetMetGluLeuAlaGlnIleTyrLysLysTyrAla 420
DB 1266 CGATGAGAGATGACGAAAGAAAGCAATGATGGACTTGCCTCAAAATTTATAAAGAAATGCT 1325
QY 421 LeuGlnSerAlaAlaGluLysAspAlaAlaLysGlnIleAlaTyrIleLysAspLysLeu 440
DB 1326 TTACAGTACGACGCTGGAAAAAGATGCTCCAAAACAGATAGCATGATCAAAAGACAAATG 1385
QY 441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
DB 1386 CTACATATATATTAACAATAATAGTATGATGATGACACTGCTTGAACGAGATCTTCT 1445
QY 461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
DB 1446 CAATACAGAGTTCCTCCAAATTTTAAACCTACAGAACGATGAATAATGCTTATTTACTTG 1505
QY 481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTyrLysCysGlnAsn 500
DB 1506 TATGCCACACTGGATTTTAAATGCTGTGAAGACATTTGAATGAATGTGCAAAAT 1565
QY 501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
DB 1566 CTGCTCCGACATCAAGTAAGAGATTTGCTTGCCTGATTAAGCAACCAAAACAGATGCC 1625
QY 521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
DB 1626 AGTGTCAAAGGCCATATTTTCAAAAGCTGATGCTTATTAACAAGACTTTCCTGATCCCTGCT 1685
QY 541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560

```
|||||
Db 1686 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTAGAGATGATGAGAAAAATAGA 1745
QY 561 LysGluLeuGluValLeuValSerProThrCysSerCysGlnAlaGluGlyCysVal 580
Db 1746 AACCAAGTTAACTACTGTTAGTCCAAACATGCTCCGCAAGCAGGCTGAAGGTGGTG 1805
QY 581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
Db 1806 CGTGAATATACATAAGAGTTGGGCAACCCAAACAGCTTCAATCTTCCCTGGAAATG 1865
QY 601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
Db 1866 ATCAAGTTTCTCTGGAGAGATAGACACCTGTGCACATAGATACCGAATCTATCAGTGC 1925
QY 621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluVal 640
Db 1926 CTTATTTAAACAATGAAACAAATCAATGATGAAACAGCATGATGAAATGAGGCTGTT 1985
QY 641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
Db 1986 CCAACGATCAACACCATCAGACAGGCTTGAACTGCTTAAGTACTCTCATTTACACAT 2045
QY 661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
Db 2046 CCCATGTCATTCATTCCTGCTGAACATTTGAATTCATTACTGGCCTGCTGAAATATGAT 2105
QY 681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
Db 2106 GATGAATAAGTACAGACAGCTGCATCAAAATTTTCAAAACACAGGAACCAAAATTTGAA 2165
QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisIleLysSerLys 720
Db 2166 GAGCATTTTCCACACATCAGATCAGCCTTCTCCGTGTTTACATCAACAAATCTAAAAA 2225
QY 721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerLys 740
Db 2226 GGACCCCCCGTCAAGCAAAATATGCCATTCATTTGATTCATCGATATTTTCTAGTAAA 2285
QY 741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db 2286 GAGGCCAGTTTCACACAGATATTGAGCCCTCTGCATAGAGCCTAGATCCAAACACCTG 2345
QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db 2346 GAACATCTCATTAACACCATGTTTACTATGTGATATGCTCTCTCCCTGCACCTGATCAA 2405
QY 781 PheAlaAlaProThrLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
Db 2406 TTTCCTGCTCTCTTGAATCTTTGGTACCTTTCATTTGTGAAGATCTTCTCATGAAAT 2465
QY 801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
Db 2466 GATCGGCTTCAGGAAAAAAGACAACTTGGGTTCCAGATGAGAAGATATCTCCT 2525
QY 821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
Db 2526 GAACACATGCTCAAAATTCAGGCTATTAATGATGATGATGATGATGATGATGATGATGAA 2585
QY 841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
Db 2586 AATAATACACAGTAAATCAGGAACCTTCACTTAAGATGCTAACAAATAATTTGCAAGT 2645
QY 861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
Db 2646 GATGAGACCTTGACAGACAGGAGGAAAAATTAGTAACACAGATATGTCACGCTGAGACCT 2705
QY 881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
Db 2706 GCTGCTGGAGTGTCTATTTGAGAGCTGGCACAGAAACCTGTTACCATGAAATCATCACA 2765
QY 901 LeuGluGlnThrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
|||||
Db 2766 TTGAACAATATACAGTATGTCATTAAGCTATATCAACGATGAATCATCAAGTAACACA 2825
QY 921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluIleGlyMet 940
Db 2826 GTGTTGGCCAGAACTTTCACAAAGGCTTTCCCGTTTACGGCTTCCACTTGAGTATAG 2885
QY 941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluIleArgAlaHisIleAlaArgGln 960
Db 2886 GCATCTGTGGCCCTTTGTGCAAAAAGATCCGTAAAGAGAGAAAGACCTCATGTAGACAA 2945
QY 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
Db 2946 TGTTTGGTGAATAATATTAATGTAAAGCCGAGATATCTCAAGCAGCATCAGCTGTTAAT 3005
QY 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3006 GAAAAATTAATGCTCTTCTTACACAGATATGTTGTTCCATATTAATTAACCTTTGGCA 3065
QY 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3066 CATGACCCAGATTAATGCAAAAGTACAGAGATATTGAACAACCTTAAGATGTTAAAGATGT 3125
QY 1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
Db 3126 CTTTGGTTTCTGTAATAATTAATGCGTAAAAAAGAAATTAACAGTCACGCTTTTATC 3185
QY 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3186 AGAAAGATGCTAATAAATATTAAAGCAACAAAGATGCCCAAGACCATGATGCAAAA 3245
QY 1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3246 ATGAATGAATAAACTGATACACTGTGTGATGTCATGAATATCATATGTCAAAGAT 3305
QY 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPheThrGln 1100
Db 3306 ACTACATACAGTTTGAATCTCTTAAAGACCCGCTACTACACACTGCTTTCTTCACTCAA 3365
QY 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Db 3366 CCTGACAGAGATTTTCAATAACACCAAAATTAATCTCTCTGATGAAATCATTTTTC 3425
QY 1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3426 ACTCTCGGAAAAACCTAAACAAACCAATGTTCTGTGAGGCTGTAAACAGCCACTTTCATCA 3485
QY 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3486 GCAGGCAAGCAATCTCAGACCAATCATCAAGATGGAACGTGAGCAATGCAAGCAGC 3545
QY 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3546 AGCTCAATTCACAGCTCTCTCTGGAAGAATTAAGAGGAGGCTGATATGTTCTGAATAGAT 3605
QY 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3606 CACAGTGAATAATTAACATTAACAATGCTTTCACCTTTCGCGGGAAGAAAGATGACAAG 3665
QY 1201 ArgAspAspSerSerLeuValArgSerGluLeuGlnLysProArgGlyLysArgLysThr 1220
Db 3666 AGAGAGCACTCTATCTTGTAAAGTCTGAAATTTGGAAGGCTTAAGAGCGAAAAAAGG 3725
QY 1221 ProValThrGluGlnGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
Db 3726 CCGCTCAGAAACAGGAGGAGGAATTAAGATGATGATGATGATGATGATGATGATGATGAA 3785
QY 1241 GluLysProLysGlySerGlnArgSerArgLysArgGlyHisIleThrAlaSerGluSerAsp 1260
Db 3786 CAGAGACCTTAAGAGCAGTACGCAAGTCCGAAAAAGGCGCATCGGCTTCAGAAATCTGAT 3845
QY 1261 GluGlnGlnThrProGluGluLysArgLeuLysGluAspIleLeuGluLysGlnGluVal 1280
Db 3846 GAACACAGTGGCTGAGGAAAAAGAGGCTCAAGAAAGATTAATTAAGAAATTAAGATGAA 3905
```

QY 1281 GlnaSerProProlysGlyLysArgGlyArgProProlysProLeuGlyGly 1300
|||||
Db 3906 CAGATAGTCCCAAGGTAAGAGAGCCACCAAACTCTTGTTGGAGGT 3965
QY 1301 ThrProlysGluGluProThrMetLysThrSerLysGlySerLysLysSerGly 1320
|||||
Db 3966 ACACCAAGAAAGAACCAATGAAACTTCTTAAAAAGGAAGCAAAAAAAATCTGGA 4025
QY 1331 ProProAlaProGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
|||||
Db 4026 CTTCCAGACAG 4085
QY 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlnSerProGlu 1360
|||||
Db 4086 TCACAAAGCAAGACAGACAGAGTGTCAAGAGAGACAGACAGAGCAAACTCTTGA 4145
QY 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
|||||
Db 4146 TCTAGTGCATTTGAATCCACACAGTCCACACAGAAAGAGAGAGAGAGAGAGAGAG 4205
QY 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
|||||
Db 4206 AGCCATCATCATCAACCAAAAAAATGTG 4238
RESULT 4
AY102267 5317 bp mRNA linear ROD 01-JUN-2003
LOCUS
DEFINITION Mus musculus androgen-induced prostate proliferative shutoff
associated protein AS3 mRNA, complete cds.
ACCESSION
VERSION AY102267.1 GI:31321922
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 5317)
Geck,P., Maffini,M., Sonnenschein,C. and Soto,A.M.
TITLE The AS3 proliferative arrest gene has an ancient eukaryotic
heritage and shows highly conserved functional domains in mice
Proc. Annu. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
JOURNAL
REFERENCE 2 (bases 1 to 5317)
Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
AUTHORS Extreme domain conservation and expression of the mouse AS3
TITLE proliferative arrest protein
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 5317)
Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
AUTHORS Direct Submission
TITLE Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
JOURNAL School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
FEATURES
source
1. 5317
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/C"
/db_xref="taxon:10090"
/chromosome="5"
/map="84cm; G2: between BRCA2 and KL"
/sex="male"
/tissue_type="brain"
147. 4487
/codon_start=1
/product="androgen-induced prostate proliferative shutoff
associated protein AS3"
/protein_id="AA052216.1"
/db_xref="GI:31321923"
/translation="MAHSKTRNDGSKIPPYGVKISPKISKEMVRRLKVVPTPKM
MDQSEPEKXYNLALHLSADFLKHPDKRLLVACCLDIFRITVAPAPYTPKM
LKDFMFTTRDLKLEDTKSPQFNRYFLKLNIAWKSINCFELEDNEFETQYRT
LFVYINGNHGVHMHVADLMSIICEDDTVSQELDLIVNLVPAHNLNKAAYDLA
KALKRTAOAIEPYITNPNVLMIGKTSISDLSERHVDLILELYNIDSHLLSVLPQ

LEFLKSNDEERLQVVKLLAKMEGAKDSBLASONKPLMOCYLCRFNDIHPILRECY
KASHCLNHNHDLKDLLEYLVKSHVDEDEALRDVYISYTAKKDILVNDHLNF
VREHTICLRMRVRKREAMGLAOIKYKSSLSAARDAKOSWPKDKLHYXONSID
DRLVERIFAOYWPVPHNLETTBRKCLLYLATLIDLANVAKLNEKRONILRRQYKD
LIDLKOPKTDASVKALFSKVMVITRNLDPDGKODEKKKFTQVLEDEKTRKOLEAL
VSPSCSKOAGCVKREITRKLGWNPORPNPRLKFLERLAVHIDTESISALIKO
VNSIDGTADEDEGVPTDOALRAGLETLKLVFTPHPIPSFSAETFEELACLMDDE
KVAEALQIOPKNTSIEDEPPHRSALPLVHLHKSRSKGPROMKVAIHCIALFEESK
ETFOAOFPEPLHKSILDPSENLEIPLVYITGHIILALADPQAAPLSLVAFYIKDL
MNDRLPGKKTTRKLAVPDEVSPETRVAKIQATIKMAYRRLGLKNNHKSQSTLELTT
ILHSDGLTTEQKLSKPDMSRLRLAAGSIVKLAQDEPCYHEITILEDOQLCALINDE
CYOYROYFAOKLHKGLSRLRLPLEYMAICALCADDPVKEREHAROCVKNWTVRREY
LKOHASVSEKLSLPLEYVVPYTHLHLDDEYKQVODIEQLKDVKECLMVTIELLA
KNENNSHAFIRKRWENIKOTKQADGPDPTKNEKLYTCVDAAMNINSKSTYSLEP
KDPYLPAREFTPOPKPNSNTKNYLPEPEKSPFTGPKPTAVILGAVNKLPSACKSO
TKSSRMETVSNASSSSNPSGCRIRKGLRDSSEMHSSEDEDTMSPLPGRKSDKREDP
DLSELEKRSRKKAPVTDPEERLGHDDLTPLKLOEOKRPGSGRGRKGTASDQOM
PEERKHKEILLNEDENONSPKRRKGRPPRLAGGTSKEEPTKSKGRKKLLVP
VVDDDEEEROIGMTEHKSRSKOHRTSKRAOORASPESTSVESSTOSTPOKGRPSK
APSPQOPPKIRVRSKOVATKENDSSEMDVLQASPVSDTTQOEGAEEDDISVGNV
RRSSKRERRR"
BASE COUNT 1676 a 1100 c 1199 g 1342 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5317
Score: 6942.00 Matches: 1344
Percent Similarity: 97.84% Conservative: 17
Best Local Similarity: 96.62% Mismatches: 28
Query Match: 96.51% Indels: 2
DB: 10 Gaps: 1
US-09-512-581b-2 (1-1391) x AY102267 (1-5317)
QY 1 MetalHisSerLysThrArgThrAsnAspGlyLysIleThrTyProProGlyValLys 20
Db 147 ATGGCTCATTTCAAGACAGACCAACGATGGGAAATTAATTAATCCCTCGAGATCAAG 206
QY 21 GluIleSerAspLysIleSerLysGluLueValArgArgIleLysMetValLys 40
Db 207 GAAATCTAGATTAATAATCTTAAAGAGAGATGGAGCGGTGTTAAAGATGGTTTAA 266
QY 41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLysGluLysLeuAsnLeuAla 60
Db 267 ACTTTCATGAGCATGACGACGACGCTCTGAAGAGAAAGAACTTATCTTAACCTACT 326
QY 61 LeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
Db 327 TTACATCTTGGCTTGCATCTTCCTCAAGCATCTGATTAAGATGTTCTGTACTGCTG 386
QY 81 AlaCysCysLeuAlaAspIlePheArgIleTyrrAlaProGluAlaProTyrrThrSerPro 100
Db 387 GCTTGTCTGCTTGCATGATTTTTCAGGATTAATGCTCTGAGGCTCTTACACGCTCC 446
QY 101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
Db 447 GATAAATCAAGAGATATTTATGTTTATTAACAAGGCACTGAAGGAGATGATCA 506
QY 121 LysSerProGlnPheAsnArgTyrrPheTyrrLeuLeuGluAsnIleAlaTrpValLysSer 140
Db 507 AAGAGCCTCAGTTTAATAGATATTTTATTTACTTGAACAATTCATGATGGTTAAATCA 566
QY 141 TyrrAsnIleCysPheGluLueGluLueAspSerAsnGluIlePheThrGlnLeuTyrrArgThr 160
Db 567 TATAACATATCTTGAAGTGAAGATGCAATGAATCACTTACCACTTACGAACA 626
QY 161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
Db 627 TTATTCATGATTAATAACAATAGCCACAAATGAGAAAGTTCAATATGACATGCTGACCTC 686
QY 181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
Db 687 ATGAGTTCTATCATTTTGAAGTGATACCTGATCTAGAGAGCTTGTGATATACAGATTT 746

QY	201	ValaIsnLeuValProIaHisLbysAsnLeuAsnLysGlnIaIaTyraSPLeuAlaLysAla	220
Db	747	GTAAATCTGGTACCTGGCCCATPAGAACTTAAACAAGCAAGATATGATGATTTGGCAAAAGCT	806
QY	221	LeuLeuLysAaGThrAlaGlnAlaIaIaGluProTyrlIeThrThrPhePheAsnIVal	240
Db	807	TTGGCTGAAGAGACTGCTCAAGCTATTGAACCATATATTACCAATTTTTTTTACCAGGTT	866
QY	241	LeuMetLeuGlyLysThrSerLLeSerAspLeuSerGlnHisValPheAspLeuIleLeu	260
Db	867	CTGATGCTGGGAAAACATCTATACAGTATTTGCTGAGCAATGCTTGATTTGATTCCTTG	926
QY	261	GluLeuTyrrAsnIleAspSerHisIleuLeuLeuSerValIleuProGlnIleuGlnPheLys	280
Db	927	GAACTCTACAAACATTGACAGCCATTGGCTGCTTCTGTGTCGCCCACTTGAGATTGATAA	986
QY	281	LeuLysSerAsnAspAsnGlnGluIaArgLeuGlnValValLysLeuLeuAlaLysMetPhe	300
Db	987	TTAAAGACCAATGATATAGAAAGACGCCCTCCAGGTGTATTAACATCTGGCCAAATGTTT	1046
QY	301	GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuThrPGLnCysTyrlLeuGly	320
Db	1047	GGGGCAAAAGGATTGAGAAATGGCTTCACAGAAACAAACCCCTGGACATGCTACTCGGCG	1106
QY	321	ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys	340
Db	1107	AGGTTTAATGATATTCACATGTACCAATACCCCTGCATATGTGAAATTTGGCAGCCATCG	1166
QY	341	LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrlLeuLysValArgSerHis	360
Db	1167	CTCATGTAAACCTCTCGATTGTAGCAAAAGATTTAACAGAGATTCCTTAAGTGAAGTCACAT	1226
QY	361	AspProGlnGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys	380
Db	1227	GATCCCTGAGGAAGCTATTGACATGATGTATTGTCTATATAGTAACAGCTGTAAAGAAG	1286
QY	381	AspIleLeuLeuValAsnAspHisIleuLeuAsnPheValArgGluArgThrLeuAspLys	400
Db	1287	GATATTCCTCTTGTCACATGATCACACTTCTCATTTTGTGACAGACCAACATTTGGACAAA	1346
QY	401	ArgThrPArgValArgLysGluAlaMetGlyLeuAlaGlnIleTyrlLysLysTyrlAla	420
Db	1347	CGGTGGAAAGATACGTAAGAAAGCCATGTATGGAGCTGGCCAGATTTATAGCAAAATACCT	1406
QY	421	LeuGlnSerAlaIaGlyLysAspAlaIaLysGlnIleAlaThrPILeLysAspLysLeu	440
Db	1407	TTACAGTACAGAGCTGGCAAGAAATGCTCAAAAACGATATCTCGGTCAAGACCAAGCTG	1466
QY	441	LeuHisIleTyrlTyrlGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla	460
Db	1467	CTGCATATATACATACAGACAGCATGACGACCGGCTACCTTGTTGACGCGATCTTTGCT	1526
QY	461	GlnTyrlMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrlTyrlLeu	480
Db	1527	CAGTACAAAGTGGCCCTCACAAATTTGGAAACGACGCAACGATGAATGCTTGATATTACTTG	1586
QY	481	TyrlAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetThrPILysCysGlnAsn	500
Db	1587	TACGCTACACAGACTTGAATGCTGTGAACACATTTGAATGAATTCGGAAATGTCATAAT	1646
QY	501	LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla	520
Db	1647	CTACTCCGACATCAAGTAAAGATTTACTCTGATTTTAATTAACAACCAACCAAAACAGATGCC	1706
QY	521	SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly	540
Db	1707	AGTGCACAGGCCATATATTTCTAAAGTGTATGTATTCACAAAGAAATTTGCCAGATCTGTGT	1766
QY	541	LysAlaGlnAspPheMetLysLysPheThrGlnValIleuGluAspAspGluLysIleArg	560
Db	1767	AAGGCCCAAGACTTCATGAAAGAGTTCCACAAAGTGTGGAAAGATGATGAGCAAAATATACGG	1826

QY	561	lystInleuValleuValSerProThrCysSerCyslyGlnAlaGlnGlyCysVal	580
Db	1827	AAAGAGCTGGAGGCCCTTTTGTAACTCCACATGTTCTCGAACAGGCTGAAGGCTGTGTG	1886
QY	581	ArgGluLeuThrIlySlySlyLeuGlyAsnProLysGlnProThrAsnProPheLeuGlnMet	600
Db	1887	CGTCAAAATTAACTAAGAAGTTGGGCAACCCCTAACACACCTTACCAACCCCTTCTTGGAAATG	1946
QY	601	IleuSlyPheLeuLeuGlnArgIleAlaProValHisIleAspThrGluSerIleSerAla	620
Db	1947	ATCAAGTTTCTCTGGAGAGGATAGCTCCCTGTACATGACATGAACTCAATCCATGCT	2006
QY	621	LeuIleuSlyGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlnGlyVal	640
Db	2007	CTTATTCAGGACAGGAAATTAATGAATGAAACAGCAGATGATGAGGATGAGGCTGT	2066
QY	641	ProThrAspGlnAlaIleArgAlaGlyLeuGlnIleLeuLysValLeuSerPheThrHis	660
Db	2067	CCAACGTATACAGCTATACAGGAGGCGCTTGAGCTGCTGAAGTGTGCTTTTACGGAC	2126
QY	661	ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp	680
Db	2127	CCATCTCTCCCTTACCTGCTGTGAGAGGTTGATGCCCTTCTGGCTTGTCTGAAGTCAGAT	2186
QY	681	AspGluLysValAlaGlnAlaIleLeuGlnIlePheLysAsnThrGlySerIlyIleGlu	700
Db	2187	GATAGAGAGGTGGCAGAGCTGCACATPACAAATTTTCTAAAACACAGGAACCAAAATTTGAA	2246
QY	701	GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisIlySerLysLys	720
Db	2247	GAGGATTTCCCTCACATCAGATAGAGGCTTGCTTCCGTGATCCATACCAAAATCCAAAAA	2306
QY	721	GlyProProArgGlnAlaLysIlyValAlaIleHisCysIleHisAlaIlePheSerSerLys	740
Db	2307	GGACACACCTGGCGCAGGCCAAATAGCCCATCCATGTATTCACAGCCATATTTCTGTGAAA	2366
QY	741	GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu	760
Db	2367	GAACCCCAAGTTGGACAGATATTGAGCCCTGTGCATAAAGCCCTGATCCAAAGCAACCTG	2426
QY	761	GluHisLeuIleThrProLeuValIleThrIleGlnHisIleAlaLeuLeuAlaProAspGln	780
Db	2427	GAGCATCTTTAAACCCCTCGTGGTACATATTGGTCAATATGGCTTCTGACACCTGATGAG	2486
QY	781	PheAlaIleAlaProIlyLysSerIlyProValAlaIleThrPheIleValLysAspLeuMetAsn	800
Db	2487	TTCCGCTCCTCCTCGTAGAGTCTTTGGTGGCAACTTCATGTGAAGACCTCCTCATGTAAT	2546
QY	801	AspArgLeuProGlyLysLysThrThrLysLeuThrValProAspGlnGluValSerPro	820
Db	2547	GACCGGCTTCCAGAAAAAAGAACACCTTAAGCTTTGGGTTCCAGATGAGAACTCCACCT	2606
QY	821	GluThrMetValLysIleGlnAlaIleLysMetMetValArgIlyProLeuGlnLysMetLys	840
Db	2607	GAGCATTTGGTCAAAATTCAGGCTATTAAATGATGTGTCGATGGCTACTTGGAATGAG	2666
QY	841	AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer	860
Db	2667	AATATACAGATTAAGTCAGAACTTCACCCCTCAGATTTACTAAACAACGATCTCATGTAGT	2726
QY	861	AspGlyLysPheThrGlnGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu	880
Db	2727	GATGGGATTTTGACAGAACAGGGAATTAAGTAAACAGATATGTCCACGCGCTGAGACTT	2786
QY	881	AlaIleAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlnIleLeuThr	900
Db	2787	GCTCTCGAGGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTATCCACAGATCATTTACA	2846
QY	901	LeuGluGlnIlyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln	920
Db	2847	CTGGAGGACAGTACCAAGCTGTGTGATTAGCCATTAATGATGAGCTGTATTACATGACGGCAG	2906
QY	921	ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGlnIlyrMet	940

```

|||||
Db 2907 GTGTCGCCAGAACTTCACAAAGCGCTTCCCGCTTACGGCTTCCCTTGAGTACATG 2966
Oy 941 AAlaIleCysAlaLeuCysAlaLysAspProValLysGluAArgAlaHisAlaArgGln 960
Db 2967 GCCATCTGTGCTCTTGTGGCCAAAGACCTGTGAAGGAGAGCGCGAGCCATGCTGTAGACG 3026
Oy 961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaValSer 980
Db 3027 TGCCTGGTGAAGAAACACACACTGTGAGGAGGAGTACCTGAAAGCACACATGCATTTAGT 3086
Oy 981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
Db 3087 GAAAAATATATGTCTCTCTACAGAGATGTGTTTCATATACAAATTCACCTTTGGCA 3146
Oy 1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
Db 3147 CATGACCCAGATTAATGTCAAAGTACAGATATTGAACAACCTTAAAGATGTGAAGAAATG 3206
Oy 1021 LeuTrpPheValIleGluIleLeuMetAlaLysAsnGluAsnSerHisAlaPheIle 1040
Db 3207 CTTGGTGTGTTCTGGAGATATTGATGGCTTAAATGAAACACACAGCCATGCATTTATC 3266
Oy 1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
Db 3267 AGAAAAATGTAAGAAATATTAAACAGACAAAGATGCTCAAGGACGAGATGATCAAAA 3326
Oy 1061 MetAsnGluLysLeuTyrThrValLysAspValAlaMetAsnIleIleMetSerLysSer 1080
Db 3327 ATGATATGAAATATGTCACCGTGTGTATGTTGCCATCAACATCATCATGCTCCAAAGAC 3386
Oy 1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
Db 3387 ACCACGTACACCTGGAGTCTCTCAAGAGACCCCGGCTCCAGCTGCTTTTTCACCCAG 3446
Oy 1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
Db 3447 CCTGACAAAGATTTAGTAACACCAAAAATTAACCTGCCAGAAATGAAATCATTTTTC 3506
Oy 1121 ThrProGluLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
Db 3507 ACTCCTGGAAACCTTAACAGCCAAATTTCTCGAGAGAGTTAAATTAACCTTTCATCA 3566
Oy 1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
Db 3567 GCAGGCAAAAGTCTCAGACCAANTCATCAAGAAATGAAACTGTGACGACCAAGCAGC 3626
Oy 1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
Db 3627 AGCTCCAAACCAACCTCTCTCGAAAGATCAAGGGAGGCTTGATAGCTCGAAATGGAT 3686
Oy 1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
Db 3687 CACAGTGAATAATGAAGATTATACATGCTTCCCTTGGCCAGGAAAAAAGTGACAG 3746
Oy 1201 ArgAspAspSerSerAspLeuValArgSerGluLeuGlnLysProArgGlyArgLysLysThr 1220
Db 3747 AGAGAAGACCCCTGATCTT-----TCTGAGTTGGAGAGAGCTTAGAAGTCGGAAGAA 3800
Oy 1221 ProValThrGluGlnGluLysLeuGlnLysMetAspAspLeuThrLysLeuValGlnGlu 1240
Db 3801 CCTGTACACAGACCTCTGAAGAAATTAAGTATGATGACTTAACCTAAGTTGTGACAGAA 3860
Oy 1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyLysIleThrAlaSerGluSerAsp 1260
Db 3861 CAGAAACCTTAAGCAGTACGAGCGAGCGGAAAAAGGCGCTACAGCTTACACTCAAC 3920
Oy 1261 GluGlnGlnTrpProGluGluLysArgLeuLysGlnAspIleLeuGlnAsnGluAspGlu 1280
Db 3921 GAGCAGCAGGTGCTCTGAGGAGACAAAGAGAGCTCTGTGAAAAATGAGAGATAG 3980
Oy 1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
|||||

```

```

Db 3981 CAGAACAGCCCAACCAAAAAGGCGAAAGAGGAGCCACCAACCAAACTCTGTGTGGGGG 4040
Oy 1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysSerGly 1320
Db 4041 ACATCGAAGGAAGAGCCCAATGAAATCAAAAGAAAGAAACAAACAAAAAATTGTA 4100
Oy 1321 ProProAlaProGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1340
Db 4101 CCTCTGTAGTAGACAGCAGATGAAAGAAAGAAACAAATTTGAAACACAGAACATTAAG 4160
Oy 1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
Db 4161 TCAAAAGCAAAACAGACCCGACATCAAAAGAGACACAAACAGAGCAGAAATCTCTGAA 4220
Oy 1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
Db 4221 ACAAGTGCAGTTGATTCACACAGTCCACACAGAAAGAGCAGGAAGACATCAAAA 4280
Oy 1381 ThrProSerProSerGlnProLysLysAsnVal 1391
Db 4281 GCACCATCACCATCAACACCCCAAAAAAATC 4313

RESULT 5
LOCUS AK122414 5944 bp mRNA linear ROD 15-MAR-2003
DEFINITION Mus musculus mRNA for mKIAA0979 protein.
ACCESSION AK122414.1 GI:28972559
VERSION AK122414.1 GI:28972559
KEYWORDS FLI-CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yusa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
2 (bases 1 to 5944)
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 259-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- 6
3'-end one pass sequencing.
FEATURES
source
Location/Qualifiers
1..5944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbx00002"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
1..5944
/gene="mKIAA0979"
<1..3577
/gene="mKIAA0979"
/note="CDS is predicted by in silico analysis. Start codon
is not identified.."
/clone_start=2
/codon_start=2
/evidence="not experimental"
/product="mKIAA0979 protein"
/protein_id="BAC55696.1"
/db_xref="GI:28972560"
/translational="FDLILEYNIIDSHLLSVLPQLEFKLSNDNEERLOVYKLLAKM

```

FGAKDSFLASQNKPLMOCYLGRENDIHVPRLECVKFAASHCLNHNPLDANLDELYLK
RSHPDEPAIRHDIYIVSVITAAKDKILLVNHDLNFEVERLDRKRWKRAKMGIAOI
YKRSLOSAGAKDAKOISWVKDLIHIYONSIDRLVERIEFAQYVPHNNETTER
MKCLYATYATLDLNAVKALNEMKCONLHROVAYLDLDRKOPDTDSVAIFSKVW
ITRNLDPGKAODEPMKFTOVLEDEKTRKOLEVALVSPCSQAECVCEVETKIKGN
PKOPTNPFLEMIKFLERIAVPHIDTESIALIKOVKSIDGTADDEDEVPDDAIR
AGLELVLSFTPHISFSAETESLACLMODEKVAEALQIFKVTGSKIEDPFIH
IRSALEPVLHNRKSKGPPROAKYAIHCIAHIFSSKETQOIEPFTKIDPSLEHL
ITPLVTIGHIALALADQFAPLKSIVATFYVKDLMDRLPGKTTFLWPDVEVSPE
TMYKIOAKIMWRLMGLKNNHSGSTRLTLTTILHSGDELEOKISKIPMSRLV
LAAGSAIVKLAOEPCYEITLIEYOICALINDECOVQVFAOKLHKLSTRLEPL
EYMAICALKADPKERRAHAROCILVANTVIRREYLQONAAVSEKLSLIPYVPT
IHLAHPDVKODIEQLKDVKECLMFVLEITLAKNENSHATRMVENIKQTKA
QGPDDTKMNEKLTVCVAMNIIINSKSTYSLESDPVLPAFETQPDKNFSNTKY
LPEPMKSEFTPGKRTANVAVNKPPLSAGKOSQTSRMETVSNASSNSPSGR
IKGRILDSSEMDENEDYTMSSPLPGKRSKREDDPDLSELEKPSRKRAVYDDEEL
GMDULTLVIOEOKRSGRGRKGRRTASDEDOQWPEKHKELELNEDEONSPKK
GKGRPRPKLGGGSKSEPTMKTKSGNKKIKVPVYDDEDEEROLGNEHKSNO
HRTSKRAQRAESPERTSAVESTOSTPOKGRGRSPKARSPQPRKIRVAGSKOVATKE
NDSSEMDVLOASSPVSDPTTQEGAEEDISVGNVRRSSKRERR"

BASE COUNT 1892 a 1151 c 1251 g 1650 t
ORIGIN

Alignment Scores:
Pred. No.: 1,84e-290 Length: 5944
Score: 5659.00 Matches: 1094
Percent Similarity: 97.71% Conservative: 16
Best Local Similarity: 96.30% Mismatches: 24
Query Match: 78.67% Indels: 2
DB: 10 Gaps: 1

US-09-512-581b-2 (1-1391) x AK122414 (1-5944)

QY 256 PheaspLeuileuGluLeuTyraSnIleaspSerHisLeuLeuSerValleuPro 275
DB 2 TTTCATTGATCTTGGAACTGTACACATTTGGCGCTTTCGTTCTGCTGCC 61
QY 276 GluleuGluPheLysLysSerAsnAspAsnGluGluArgLeuGluValLysLeu 295
DB 62 CAGCTGAGTTAAATTAAAGAGCAATGATATGAGAACGCCCTCCAGGTGTTAAACTA 121
QY 296 LeuAlaLysMetPheGluAlaLysAspSerGluLeuAlaSerGluAsnLysProLeuTyr 315
DB 122 CTGGCGAAAGATTGGGGCAAGAGATTCAAGATTGGCTCTTCCGAAACAAACCCCTGG 181
QY 316 GlncystTyrlLeuGluArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335
DB 182 CAGTGCATACCTGGGACAGGTTTAATGATATCCATTCACCAATACGCCCTGGATGTGTGAAA 241
QY 336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeu 355
DB 242 TTTCGCCAGCCTGCTCATGAAACCATTCGATTTACCAAGATTAAACAGACTATCTT 301
QY 356 LysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal 375
DB 302 AAAGTGAAGTCACATGATCCCTGAGGAAGCTATTAGACATGATGTATTGTGTCTATAGTA 361
QY 376 ThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnAspPheValArgGlu 395
DB 362 ACAGCTGCTAAAAAGGATATCTCTTGTCTCAATGATACCTTACCATTTTGTGTAGAGAG 421
QY 396 ArgThrLeuAspLysArgTyrArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle 415
DB 422 CGAACATTGGACAAACGCTGAGAGTACGTAAAGAACCATGATGAGGACTGGCCAGATT 481
QY 416 TyrTyrLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTyr 435
DB 482 TATAGAAATTAATCTTACAGTCCAGAGCTGGCAAAATGCTGTGAAAAACATATCTGG 541
QY 436 IleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuVal 455
DB 542 GTCAAGAGCAAGCTGCTGATATTAATATCAGAAACAGCATCGACGACCGGCTACTTGT 601

QY 456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
DB 602 GAGGAGATCTTGTGCTCAGTACATAGGCCCTCACAAATTGGAAACGAGCAACGATGAAG 661
QY 476 CysLeuTyrTyrLeuTyrAlaThrPheAspLeuAsnAlaValLysAlaLeuAsnGluMet 495
DB 662 TGCCTGATTAATCTTGAAGCTGTACACCTGCAATGCTGTGAAGCATTTGAATGAATG 721
QY 496 TyrLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
DB 722 TGGAAATGTCAAAATCTACTCCGACATCAAGTAAAGATTACTGTGATTTAATTAACAA 761
QY 516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsn 535
DB 782 CCAAAAACAGATGCCAGCTGTCAAGGCCATTTTCTTAACTGATGTTATCACACAGAAAT 841
QY 536 LeuProAspProGluLysAlaGlnAspPheMetLysLysPheThrGlnValLysGluAsp 555
DB 842 TTGCAGATCTCTGTTAAGGCCCAAGACTTCATGAGAGTTCAACACAGTGTGGAAGAT 901
QY 556 AspGluLysIleArgLysGlnLeuGluValLysLeuValSerProThrCysSerCysLysGln 575
DB 902 GATGAGAAADPACGGAACAGCTGGAAAGCCCTGTGTAGTCCACATGTTCTCTCAAGCAG 961
QY 576 AlaGluGlyCysValArgGluIleThrLysLysLeuGluLysAspProLysGlnProThrAsn 595
DB 962 GCTGAAGCGCTGTGCTGCAATTAACCTAAAGTTGGGCAACCTTAAGCAGCTTACCAAC 1021
QY 596 ProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
DB 1022 CTTTCTTGGAAATGATCAAGTTTCTTGTGAGAGATACCTCTGTACACATAGACACT 1081
QY 616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635
DB 1082 GAATCCATCAGTGTCTTATCAAGCAAGTGAATTAATCAATAGTGAACACAGCATGAT 1141
QY 636 GluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuVal 655
DB 1142 GAGATGAGGCTGTCCACATGATCAAGCTATCAGGGCAGCCCTTGAGCTGCGAAGGTG 1201
QY 656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuAla 675
DB 1202 CTGCTTTTACGACCTATCTCTTCACTCTGCTAGACGTTGAGTCCCTTCTGCGCT 1261
QY 676 CysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
DB 1262 TGCTGAAAGATGATGATGAGAGGTCGCAAGGCTCCTACTACAAATTTCCAAACACAC 1321
QY 696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
DB 1322 GGAAGCAAAATTTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTCTGTACTACAT 1381
QY 716 HisLysSerLysGluProProArgGlnAlaLysTyrAlaIleHisCysIleHisAla 735
DB 1382 CACAAATCCAAAAAGAACCACTCGGACAGGCCAAATACCCATTCATTTATATTCAGGCC 1441
QY 736 IlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeu 755
DB 1442 ATATTTTCTAGTAAAGAAACCAAGTTTGACACAGATATTAGGCTCGCATTAAGGCTA 1501
QY 756 AspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeu 775
DB 1502 GATCCAAAGCAACCTGGAGCATCTTATTAACCCCTGCTGCTACTATTTGGTCAATATGCTCTT 1561
QY 776 LeuAlaProAspGlnPheAlaIleProTyrLysSerTyrAlaThrPheIleValLys 795
DB 1562 CTGCACCTGATCACTGCTGCTCTGCTGAAAGCTTGTGTCGAACATTTCAATGTGAAG 1621
QY 796 AspLeuLeuMetAsnAspArgLeuProGluLysLysThrThrLysLeuTyrValProAsp 815
DB 1622 GACTCTCTCATGATGACGCGCTTCCAGGAAAAAGCAACATAGCTTGGTTCAGAT 1681
QY 816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835

```

Db      1682 GAGGAAGCTCAGCTGAGACATGGTCAAAATTCAGGCGATTAAATGATGGTTGCATGG 1741
Qy      836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThreuarGLeuLeuThr 855
Db      1742 CTACTTGGAAATCAAAATATATCATGAGTCAAGTCAAGCACTTCCACCCTCAGATTACTAAC 1801
Qy      856 ThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspMet 875
Db      1802 ACGATACCTGATAGTATGGATGGGATTTGACAGAACAGGAAATTAATGTAACCAAGATATG 1861
Qy      876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGlnProCysTyr 895
Db      1862 TCACCCCTGAGACTTGTCTGCTGGAGTCTATTGTGAAGCTGGACAGAGAGCCCTGTATC 1921
Qy      896 HisGluIleIleThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 915
Db      1922 CACGAGATCTATACACTGAGACAGTACAGCTGTGTGATGAGCCATCATGATGATGCT 1981
Qy      916 TyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu 935
Db      1982 TATCAAGTCAGGACAGGTGTGTGCTCAGAACTTCACAAAGGCCCTTCCCGCTTACGGCTT 2041
Qy      936 ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 955
Db      2042 CCCCTTGAATGATGATGGCATCTGTCTCTTGTGCCAAAGACCCCTGAAAGAGAGCCGA 2101
Qy      956 AlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
Db      2102 GCCCATGTCTAGACAGCTGTGTGTGTAAGAACACACTGCTGAGAGGAGACTCTGAAAGCAG 2161
Qy      976 HisAlaIleValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThr 995
Db      2162 CATGACGCTGTATGAGAAAATATATGTCTCTCTTACCAAGACTATGTTGTTCCATATACA 2221
Qy      996 IleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLys 1015
Db      2222 ATTCACCTTTGGGCATGACACCCAGATTAATGCAAAATGACAGATTAATGCAAACTTAAA 2281
Qy      1016 AspValLysGluCysLeuThrPheValLeuGlnIleLeuMetAlaLysAsnGluAsn 1035
Db      2282 GATGGAAAGAAATGCTTTGGTTGTCTGTGAGATTAATGATGCTTAAATAAATAAACAC 2341
Qy      1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
Db      2342 AGCCATGCTATTATGAGAAAATGCTAGAAAATATTAACACAGCAAAAGATGCTCAAGA 2401
Qy      1056 ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
Db      2402 CCAGATGATACAAAATGAATGAATAATTTGTACACCGTGTGATGTTGCCATGAACATC 2461
Qy      1076 IleMetSerLysSerThrThrTyrSerLeuGlnSerProLysAspProValLeuProAla 1095
Db      2462 ATCAATGTCACAGACACACGATACGCTGTGAGTCTCTTAAGGACCCGCTGTGACACT 2521
Qy      1096 ArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProGlu 1115
Db      2522 CGGTTTTCACCCAGCCCTGACAGAAATTTTGTAGTAACACCAAAAATTTACTGCTTCAAA 2581
Qy      1116 MetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsn 1135
Db      2582 ATGAATATCATTTTTCACCTCTGAGAAACCTTAACACAGCAATGTTCTCGAGACAGTTAAT 2641
Qy      1136 LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal 1155
Db      2642 AAGCCACTTTCATATCAGCAGCAAAACAGTCTCAGACCAATCATCAAGAAATGAGAACTGTG 2701
Qy      1156 SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAsp 1175
Db      2702 AGCAACGCAAGCAGAGCTCAACCCAGCTCTCTGAGAGATCAAGGAGGAGGTTTAT 2761
Qy      1176 SerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlu 1195

```

```

Db      2762 AGCTCTGAATGATACACAGTGAATAATGAAATATATACAAATGCTTTCACCTTGGCAGGA 2821
Qy      1136 LysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlnLysProArg 1215
Db      2822 AAAAAAAGTGAACAGAGAGAGACCCGTATCTT-----TCTGAGTTGGAGAAAGCCTTGA 2875
Qy      1216 GlyArgLysLysThrProValThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1235
Db      2876 ACTCGAAAAAAGCACTGTCTACACAGCCCTGAGAGAAATTAATGATGATGACTTAAT 2935
Qy      1236 LysLeuValGlnGlnGlnLysProLysGlySerGlnArgSerArgLysArgLysIleThr 1255
Db      2936 AAGTGTGATCAGAGACAGAAACCTTAAGCAAGTCAAGCAGGCGGAGAAAGAGCCGTACA 2995
Qy      1256 AlaSerGluSerAspGlnGlnGlnThrProGluGlnGlnGlnGlnGlnGlnGlnGlnGln 1275
Db      2996 GCTTCAGACTCAGACAGACAGCAGAGGCGCTGTGAGAGAGAGCAGACAAAGAGAGCTCTG 3055
Qy      1276 GluAsnGluAspGlnGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys 1295
Db      3056 GAAATATGAGATGACACAGACAGCCACCAAAAAGGCCAAAGAGCAGCCACCAAAA 3115
Qy      1296 ProLeuGlyGlyGlyThrProLysGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1315
Db      3116 CCTCTGTGGTGGGGGACATGCAAGAGAGGCGCCACAAATGAAACATCCAAAGAAAGCA 3175
Qy      1316 LysLysLysSerGlyProProAlaProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1335
Db      3176 AAGAAAAAACTTGTACCTCTGTACTAACGACGATGAAAGAAAGAAAGCAAAATTTGGA 3235
Qy      1336 AsnThrGlnGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
Db      3236 AACACAGAACATTAAGTCAAAAAGCAAGCAGCCACATCAAGAGAGCAGCAGACAGGA 3295
Qy      1356 AlaGluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArg 1375
Db      3296 GCAGAAATCTCTGAAACAGTCAAGTCAATTCACACAGTCCACACACAGAAAGAGAGA 3355
Qy      1376 GlyArgProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
Db      3356 GGAAGACCATCAAAAAGCACATCCACATCAACACCCCAAAAATATC 3403

RESULN 6
AF294791      6744 bp      mRNA      linear      PRI 24-JUL-2002
LOCUS      AF294791
DEFINITION      Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION      AF294791
VERSION      AF294791.1 GI:21951801
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
TITLE      Identification and characterization of SCC-112, a novel cell cycle
            regulated gene in human cancer cells
JOURNAL      unpublished
REFERENCE     2 (bases 1 to 6744)
AUTHORS      Kumar,D. and Kasid,U.
DIRECT SUBMISSION
JOURNAL      Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
            Washington, DC 20007, USA
FEATURES
source
1..6744
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="4"
1..6744
/gene="SCC-112"
232..4125
/gene="SCC-112"
CDS

```


Db 1720 GGAAGACGATGACATAGCAAGAAATTTGCTGACCCCGGGAAGACACAGATTGTCG 1779
 QY LysLysPheThrGlnValLeuGlnLysAspArgLysLysIleArgLysGlnLeuGlnValLeu 566
 Db 1780 AAGAAATTTAACCAAGGTTCTCGCGATGATGAGAAACCTTGGCTCAGTTGAGATTATTA 1839
 QY 567 ValSerProThrCysSerCysLysGlnIleArgLysValArgGlnIleThrLysLys 586
 Db 1840 ATTACCCACACTGTTCTTGCAACACAGAGATATTGTGTGAGAGAAATACCCGGAAA 1899
 QY LeuGlnAsnProLysGlnProThrAsnProPheLeuGlnMetIleLysPheLeuLeuGln 606
 Db 1900 CTTCGAACATCCTAACCAACCAACAATCCTTTCTAGAGATGGTCAAAATTTCTGTGGAA 1959
 QY ArgIleAlaProValHisIleAspThrGlnUserIleSerAlaLeuIleLysGlnValAsn 626
 Db 1960 AGAATCGACCTGTGCACATTGATTCAGAGAGCCATAGCCACTAGTGAATTTGATGAT 2019
 QY LysSerIleAspGlyThrAlaAspArgLysAspGlnGlyValProThrAspGlnAlaIle 646
 Db 2020 AAGTCATAGAGGGGACAGACAGATATGATGAGAGAGGGGTGTAAGTCCAGATACAGCTATC 2079
 QY ArgAlaGlyLeuGlnLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
 Db 2080 CGTTCAGGACTTGAACTTCTTAAGGTCTGTCTTTTACACATCTACCTCGTTCACATCT 2139
 QY AlaGlnThrPheGlnSerLeuLeuLysLeuLysMetAspAspArgLysValAlaGln 686
 Db 2140 GCAGAGACATATAGATTCCTTGTACAGTCCCTAAGAAATGAGAGATGACAGGTAGAGAA 2199
 QY AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlnGlnAspPheProHisIle 706
 Db 2200 GCTGCTATTCAAATTTTGAATATACAGTCAACAATAGAAACAGACCTTCCACAGATA 2259
 QY ArgSerAlaLeuLeuProValLeuHisHisLysSerLysGlyProProArgGlnAla 726
 Db 2260 CGATGACCTTAATTTCCATTTTACATTTTACATAAAGCAAGAGGGGTACTCCACCAAGCA 2319
 QY LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGlyThrGlnPheAlaGln 746
 Db 2320 AAACAGGCTGTGCATGTATACAGCCCATATTACAAATTAAGAAAGTCCAGCTTGACAG 2379
 QY IlePheGlnProLeuHisLysSerLeuAspProSerAsnLeuGlnHisLeuIleThrPro 766
 Db 2380 ATTTTGAAGCCACTGAGTAGAGTGTGAATGCTGATGTGCCAGAACACTTATTACTCCA 2439
 QY LeuValIleThrIleGlnHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProThrLys 786
 Db 2440 TTAGTTTCAATGGGCCACATTTCTATGTTAGCACCAAGATCAGTTTGTCCCAATGAAA 2499
 QY SerTyrValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
 Db 2500 TCTGTAGTAGCAAAATTTTATTTGTGAAAGTCTGCTAATGATGACAGGCAACAGGTGAA 2559
 QY LysThrThrLysLeuThrPvalProAspGlnGlnValSerProGlyIleThrMetValLysIle 826
 Db 2560 AAGATGGAAGAACTGTGTCTCAGATGAGAGGTTTCCCTCGAAGTACTACCAAGGTA 2619
 QY GlnAlaIleLysMetLeuValArgThrLeuLeuGlnMetLysAsnAsnHisSerLysSer 846
 Db 2620 CAGGCAATTTAACTTCTGTAGAGTGTGTGGGTATGAAAAACACAGCTTAATCT 2679
 QY GlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGln 866
 Db 2680 GCCAATTACACCTTCGGTTATTATACAGGATGTTGGTTAGTGAAGGGTACCTGACAGAG 2739
 QY GlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIle 886
 Db 2740 CAAAGAGAGATCAGTAATTCGATATGTCGCTTGCATTTAGCTGCGTGATGCCATA 2799
 QY ValLysLeuAlaGlnGlnProCysTyrHisGlnIleIleThrLeuGlnGlnIleThrLeu 906
 Db 2800 ATGAGCTTGTCTCAGGAACCTTGTACATGAAATTATTACCCAGAACAGTGTTCAGCTC 2859

QY CysAlaLeuAlaIleAsnAspArgLysCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
 Db 2860 TGTGCACCTTGTATTATTAATGATAGTGTATCCAAATGAGGACATATTTGCTCAGAACTG 2919
 QY HisLysGlyLeuSerArgLeuArgLeuProLeuGlnGlyTyrMetAlaIleCysAlaLeuCys 946
 Db 2920 CATTAAGGCACTGTGAAGTTACTGCTCCCATTTGAGAGTATATGCGCATCTTTCGTTGCT 2979
 QY AlaLysAspProValLysGlnArgAlaAlaHisAlaIleArgGlnCysLeuValLysAsnIle 966
 Db 2980 GCCAAAGATCCCTGTAGAGAGAGAGACACACACGACACGACATGTTTACTGAAAAATATC 3039
 QY AsnValArgArgGlnTyrLysLysIleAlaIleValSerGlnLysLeuLeuSerLeu 986
 Db 3040 AGTATACGAGGGAATACATTTAAGCAATCTTATGCTAGCTACTAGCAAAATTTATTCACAG 3099
 QY LeuProGlyTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
 Db 3100 TTGCGTAATATGATGTTCCATACATGATTCACCTCGTATAGGCCATGATCCAGATTTTACA 3159
 QY LysValGlnAspIleGlnGlnLeuLysAspValLysGlnCysLeuThrPheValLeuGln 1026
 Db 3160 AGATCACAAGATGTGATCAGCTCTGTATATCAAAAGAGTGCCTATGCTCATGCTTGA 3219
 QY IleLeuMetAlaLysAsnGlnLysAsnSerHisAlaPheIleArgLysMetValGlnAsn 1046
 Db 3220 GTTTAATGACAAAGATGAAACAAATAGCCATGCTGCTTATATGAAAGATGCGAGAGAAC 3279
 QY IleLysGlnThrLysAspAlaGlnIleProAspAspAlaLysMetAsnGlnLysLeuTyr 1066
 Db 3280 ATCAAGTTAATCCAGAGATGCCAGCTCCAGATGATTCACCAAGACAAAGAAACACTGTAT 3339
 QY ThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGln 1086
 Db 3340 ACAGTATGATGTGCTCTGTGTATTAAATATGATAAAGCTTTGTGCATGCATCAT 3399
 QY SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSer 1106
 Db 3400 TCACCAAGAGACCCAGCTCTCCCATGAAATTTTATACACAACTGAAAGAGACTTTCGT 3459
 QY AsnThrLysAsnTyrLeuProProGlnLysSerPhePheThrProGlyLysProLys 1126
 Db 3460 AACGATTAAGATTAATTTACAGAAAGACAAAGACTTCTGTATTAACAGGAAGCCAAAG 3519
 QY ThrThrAsnValLeuGlnValAlaHisLysProLeuSerSerAlaGlyLysGlnSerGln 1146
 Db 3520 CCTGCTGAGATGATAGGTGCACTAATTAAGCTTTATACAGCAAGGAGAAACCTTAT 3579
 QY ThrLysSerSerArgMetGlnThrValSer-----AsnAlaSerSerSerSerPro 1164
 Db 3580 GTTAAGACACTGACACTGAGACTGAGCAATTAATTAATGTAATTTACAGCTGAAACCT 3639
 QY SerSerProGlyArgIleLysGlyArg---LeuAspSerSerGlnMetAspHisSerGln 1183
 Db 3640 TCAACCGGAATAATCGATCAAGGAGACAGATTTCAGAGCGACAGAACTGAGTGTGAA 3699
 QY AsnGlnAsp-----TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg 1201
 Db 3700 AATGAAGAGAACCTGTGAGGATTTATTTCTACGTACACACTGTAAAGAAATATTGAC----- 3753
 QY AspAspSerAspLeuValArgSerGlnLeuGlnLysProArgGlyArgLysLysThrPro 1221
 Db 3754 -----CCA 3756
 QY ValThrGlnGlnGlnGlnLysLeuGlnLysMetAspAspLeuThrLysLeuValGlnGln 1241
 Db 3757 GTAAGGAATTAAGAA-----ATTAATTTCTGATCAGCTAC-----CAG 3795
 QY LysProLysGlySerGlnArgSerArgLysArgGlyHisIleThrAlaSerGlnSerAspLys 1261
 Db 3796 GGCACACATCAGCAGTGCAGGAGAAAGAAAGAACATTAACAGCAGCTGCTGTCAGAGAAAT 3855

Oy	1262	---	glnGlnTrpProGluGluLysArgLeuLysLunSplLeuGluAsnGlu	1280
Db	3856	ATCCACACAAAACAGATGAGAAA-----	-----GTAGATGAA	3888
Oy	1281	GlnAsnSerPro---	ProLysLysGlyLysArgLysArgProPolysProLeuGly---	1298
Db	3889	TCGGGACACTCCCGCCCTCCCAAAACCCAGAGAGAGCTGCAGCCACGTCGATCTCAG	3948	
Oy	1239	GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLys	1318	
Db	3949	GGCAATGCTACCAAAAATGATGATCTAAATTAAACCTATTAAAGGAGAGAGAGAGCT	4008	
Oy	1319	ser---	glyProAlaProGluGluGlu-----	1328
Db	4009	GCATGGGTCCAGAGAGAGCCCTGGGGGTTTGGACACAGTAAATGCCAAAGCACCAACTG	4068	
Oy	1329	GluGluGluArgLysSerGlyAsnThrGluGlnLys-----	-----SerLysSerLys	1344
Db	4069	CAGATATT-AGCCAAAAAGCGCAGCAGCAGCAAGAAATGACTTACAAAGGTAA	4127	
Oy	1345	GlnHis-----	ArgValSerArgArgAlaGlnGlnArgLysArgLysSerProGluSer	1362
Db	4128	ATGCATTTCGCAAGAGGAGAAAATAGAGCCAAACAGAG--	CAGGCTCCAGCTTCTGCA	4184
RESULT 7				
HSU50533			1852 bp	mRNA
LOCUS				Linear
DEFINITION				PRI 27-NOV-1996
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PubMed				
REFERENCE				
AUTHORS				
JOURNAL				
FEATURES				
SOURCE				
BASE COUNT				
ORIGIN				
Alignment Scores:				
Pred. NO.:				
Score:				
Percent Similarity:				
Best Local Similarity:				
Query Match:				

DB:	9	Gaps:	1
US-09-512-581B-2 (1-1391) x HSU50533 (1-1852)			
QY	356	LysValAlaSerSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal	375
Db	3	AAAGGAGAGTCACATGACCCTGAGAGAACTTTAGACATGATGTTATGTGTCATAGTT	62
QY	376	ThrAlaAlaLysLysAspIleIleuLeuValAsnAspHisLeuLeuAsnPheValArgGlu	395
Db	63	ACAGTGTGTAAGGATATTCCTTGCTCAATGATCTCACTTAATTTGTGACAGAG	122
QY	396	ArgThrIleuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle	415
Db	123	AGACATTATGACAAACGATGGAGAGTACGCCAAGAACCATGATGGACCTGCCCAATT	182
QY	416	TyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrp	435
Db	183	TATAAGAAATATGCTTTACAGTCCAGCAGCTGGAAAGATGCTGCAAAAACAGATAGCATGG	242
QY	436	IleLysAspLysIleuLeuHisIleTyrTrpGlnSerIleAspAspArgLeuLeuVal	455
Db	243	ATCAAGACAAATGCTCAATATATATTAATAATATATGATGATGACACTGCTGT	302
QY	456	GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluTrpThrGluArgMetLys	475
Db	303	GAAGGATCTTTGGTCATACATGAGTTCCTCCACAATTAGAAGCTACAGAACGGATGAA	362
QY	476	CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMet	495
Db	363	TGCTTATATTTACTGTATGACACACTGGATTAAATGCTGGAAGCATTTCAATGAAGATG	422
QY	496	TrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuLysGlnIleLysGln	515
Db	423	TGGAATGTCAAAATCTGCTCCGACATCAAGTAAAGATTTTGCTTGACTTATTAAGCAA	482
QY	516	ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsn	535
Db	483	CCCAAAACAGATGCCAGTGCACAGCCCTATTTTCAAAAGCATGGTATTACAAAGAAAT	542
QY	536	LeuProAspProGluLysAlaGlnAspPheMetLysLysPheTrpGlnValLeuGlnAsp	555
Db	543	TTACTGTACTCTGGTAAGGCTCAGGATTTCTGGAAGAAATTCAACAGGTGTGGAAGAT	602
QY	556	AspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGln	575
Db	603	GATGAGAAATAAGAAACACAGTTGGAATACTTGTAGTCCAAACATGCTCCTGCAACAG	662
QY	576	AlaGluGluCysValArgGluIleThrLysLysLeuGluAsnProLysGlnProThrAsp	595
Db	663	GCTGAGAGTTGTGCGTGAATACCTAGAGAGTTGGCAACCCCAACACCCCTACCAAT	722
QY	596	ProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr	615
Db	723	CCTTTCCGGAATGATCAATTTCTCTTGGAAGAGATAGACACCTGTGCACATGATACC	782
QY	616	GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp	635
Db	783	GAATCTATACAGTGTCTTATTAAACAATGAACAATCAATAATGAGACGACGATGAT	842
QY	636	GluAspGluGluValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuVal	655
Db	843	GAAGATGAGGGTGTCCAACTGATCAAGCCATCAGACAGGCTTTGAACCTCTTAAGTA	902
QY	656	LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAla	675
Db	903	CTTCATTTACACATCCCATCTCATTTTCATTTCTGCTGCAACATTTGAAATCATTTCTGCT	962
QY	676	CysLeuLysMetAspAspGluLysValAlaGluAlaIleLeuGlnIlePheLysAsnThr	695
Db	963	TGTCTGAAATAAGATGATGAATAAGTACCAAGACCTGTGACTACAAATTTTCAAAACACA	1022
QY	696	GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValIleHis	715

```

|||||
Db 1023 GGAAGCAAAATTTGAAAGAGATTTTCACACATCATGACCTTCCTTCGTTTACAT 1082
Oy 716 HistySerLysLysGlyProProArg-GlnAlaLysTyrAlaIleHisCysIleHisAl 735
Db 1083 CACAACATCTAAAGGACCCGCCGCCCAAGCCAAATATCCATTCATTCATCCAGC 1142
Oy 735 aIlePheSerLysGlyThrGlnPheAlaGlnIlePheGlnProLeuHisLysSerLe 755
Db 1143 GATATTTTCTAGTAAAGAGACATGCTTTCACAGATATTTGAGCCTTCGATTAAGAGCT 1202
Oy 755 uasProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLe 775
Db 1203 AGATCCAGCAACCTCGAACAATCATATACACATATGTTACTATGTCATATATGCTCT 1262
Oy 775 uLeuAlaProAsnGlnPheAlaIleAlaIleProTyrLysSerTyrValAlaThrPheIleVally 795
Db 1263 CCTTGACACCTGATCAATTTGCTGCTCTTGAATCTTTGGATGCTTTCATTCATGAGA 1322
Oy 795 sasPLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTyrValProAs 815
Db 1323 AGATCTTCTCATGATGATGATCGGCTTCAGGCAAAAAGCAACTAATCTTGGGTTCCAGA 1382
Oy 815 pGluGlnValSerProGlnThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835
Db 1383 TGAAGAAGTATCTCCTGAGACAAATGTCATAAATTCAGGCTATTAATGATGTTGCATG 1442
Oy 835 pLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeu-ArgLeuLeuT 855
Db 1443 GCTATCTGGCATGTAAMTTAATTCACAGTATATCAGGATCTTCTCTTAAGSATTCGTA 1502
Oy 855 hTThrIleLeuHisSerAspGlyAspLeuThrGlnGlnGlyLysIleSerLysProAspM 875
Db 1503 CACAATATATCCATTCGATGCGGATTTGGAGAACAGGCGGAAT---TAGTAACCAAGTA 1559
Oy 875 eLysArgLeuArgLeuAlaIleGlySerAlaIleValLysLeuAlaGlnGlnProCysT 895
Db 1560 TGTCACGCTGAGACTTCCTGCTGAGAGTGCATATRTAGAGCTGGCACAGACCTGTT 1619
Oy 895 yThIleGlnIleIleThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 915
Db 1620 ACCATGGAATCTCAGACATAGACAAATCATGCTATGCTTACCTATCAACGATGTAAT 1679
Oy 915 ySTyGlnValArgGlnValPhe-AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg 934
Db 1680 GCTATCAAGTAAGCAAGCTGTTTGGCCAGAAACTTCGCAAAAGGCTTTCCTGTTGCGG 1739
Oy 935 LeuProLeuGlnTyrMetAlaIleCysAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnG 954
Db 1740 CTTCCACTTGAGAAATATGCAATCTGTCCTTGTGCAAAAAGACCTGTAAGAGAGAGA 1799
Oy 955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGln 971
Db 1800 AGAGCTCATGCTAGGCAATGTTTGGTGAATAATATTAATGAAGCGGAGAC 1850

```

```

MEDLINE 98403880
PUBMED 9734811
REFERENCE 2 (bases 1 to 5177)
AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail: cdna@info.kazusa.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

FEATURES
Source
1..5177
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="H303994"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1..5177
/gene="KIAA0648"
<1..2558
/gene="KIAA0648"
/codon_start=3
/product="KIAA0648 protein"
/protein_id="BAA31623.1"
/db_xref="GI:3327110"
/translation="LTVLYASLDPRNAKALNDENKCOMLRSHVRELLDKHQPSEA
NCSAMFRLKMTIARNLPDPRGAODPVKKFENVOLDDEKRLSRLQLELLSPSCROADI
CVREIARLKLMPKPTNPFLEKRVLEKRIKADPKYERARAROCLENNISIREYIKONPMATEKL
EGEVSPDPAIRSGLELKVLSFTPTSPSHAEYSELSQCLMEDDKVAEAIQIFRN
TGHKIEDLPQIRSLIPILIHOKAKRGPRHAKAVHCHIAIFNKVEQLQIDFPLS
RSINADVPQOLITPLVSGHTSMILAPDPAIPMKSVANFTYKDLMLNDSRSTGRKNC
LWSPDEKSPPYLAKVQAIKILYRLKLCMKRNOXSANSTRILSAMLVSGCDITBOK
RISKSDMSRLRLAAGSALIMKLAQECHEIITPEQPCALVINDECYQVQITRQKL
HKALVKLLPLEYMAIFALCAKDKPYERARAROCLENNISIREYIKONPMATEKL
SLPEYVYPMIHLIADPDPTSDVDQDLDEKLEMLFLEVLMTKNNNSHAFMK
MAENIKILRDOASPDSEKTNKEIYCDVALCVINRSKALCNADSPKDPVLPMPFPO
PEKPCNDKSYISERTRYLLITGKPKPGVIGAVNKPISATGRKPYVSTGTGFSNI
NVNSLNPSTNRSRBOSSSEAEVGSSENEENPVRIISVTVKNIIDPVKNEINSDOA
TQGNITSSDRGKRRTVYTAGAENIQDKTEKVEDSGPPAPSKPRGRKRSISQVATK
NDDLNKTPINKRKRRAVGOESPGLGLENARAPKLDIAKRAAPAEIROIDLQR"
BASE COUNT 1690 a 931 c 1037 g 1519 t
ORIGIN
Alignment Scores:
Pred. No.: 5,26e-134 Length: 5177
Score: 2704.50 Matches: 551
Percent Similarity: 74.07% Conservative: 123
Best Local Similarity: 60.55% Mismatches: 174
Query Match: 37.60% Indels: 63
DB: 9 Gaps: 15
US-09-512-581b-2 (1-1391) x AB014548 (1-5177)
Oy 477 LeuTyrTyrLeuTyrAlaThrLeuAsnLeuAsnAlaValLysAlaLeuAsnGlnMetTyr 496
Db 3 TTATATTAATCTATATAGCTAGATTGATCCAAATGCTGTGAAGGCTCTCAAGCAAAATGAGG 62
Oy 497 LysCysGlnAsnLeuLeuArgHisGlnAlaValLysAspLeuLeuAspLeuIleLysGlnPro 516
Db 63 AAGTGTCAAGACATGCTTCGAGCAGCATATGACGCAACATTTGATTCACACAGAGAGCT 122
Oy 517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
Db 123 ACATCAGAGCGCTAATGTTCTGCAATTTGTGGAACACGATGACCATCAAGCAAGATATT 182
Oy 537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGlnLysAsp 556
Db 183 CTTGACCCCGGAAAGCACAAAGATTTTGTGAAGAAATTTAACCAAGGTTCTGCGCGATCAT 242
Oy 557 GlnLysIleArgLysGlnLeuGlnValLeuValSerProThrCysSerCysLysGlnAla 576
|||||

```

JOURNAL

Db 243 GAGAACTTCGCTCAGTTGGATTATTAATTAAGCCCAACCTGTTCTTGCAAAACAGCA 302
QY 577 GUGUGYcysValargIuilethrlYsLysLeuGlyYasProLysGlnProthrasPro 596
Db 303 GATATTGTGTGAGAAATAGCCGGAACCTTGCAATCTTAAGCAACCAACAAATCCT 362
QY 597 PheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGlu 616
Db 363 TTTCGAGATGCTCAAAATTTCTGTGGAAAGAATGCACCTGTGCACATGTGATTCGAA 422
QY 617 SerIleSerAlaLeuIleLysGlnValAspLysSerIleAspGlyThrAlaAspArgIu 636
Db 423 GCATAGTGCCTAGTGAATGAAATTAAGTCAATAGAGGGACAGAGATGATGAA 482
QY 637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuValLeu 656
Db 483 GAGAGGGTGTAGTCCAGATACAGCTATCCGTTCCAGACTTGACCTCTTAAGGTTCTG 542
QY 657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
Db 543 TCTTTACACATCCTACCTGCTTCCACTCTGCAGACATATGAGTCTGTGACAGTGC 602
QY 677 LeuLysMetAspArgIuLysValAlaGluAlaLeuGlnIlePheLysAsnThrGly 696
Db 603 CTAGAGATGAGATGCAAGGTAGCAAGAGCTGATTCGAAATTTTAAAGAAATACAGGT 662
QY 697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
Db 663 CACAAATATGAACACAGCTTCCCGCAGATTCGATCCAGCTTAATTCCTCATTTTACATCA 722
QY 717 LysSerLysGlyProProArgGlnAlaLysThrAlaIleHisCysIleHisAlaIle 736
Db 723 AAGCAAGAGGGGTACTCCACACCAAGCAAAACAGCTGTGCACGTATACAGCCATA 782
QY 737 PheSerSerLysGlnThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
Db 783 TTACAAATTAAGAGCTGCTGCACAGATTTTGGAGCCACTGTGACAGAGCTGTAAT 842
QY 757 ProSerAsnLeuGlnHisLeuIleThrProLeuValThrIleGlnHisIleAlaLeu 776
Db 843 GGTGATGTCCAGACACACTTATACCTCATAGTTTCATGTGGCCACATTTCTGATGTA 902
QY 777 AlaProAspGlnPheAlaAlaProThrLysSerThrValAlaThrPheIleValLysAsp 796
Db 903 GCACCAAGTACAGTTGTCTCCCAATGAATCTGTAGTACAAATTTATTTGCAAGAT 962
QY 797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuThrProAspGlu 816
Db 963 CTCTAATGATACAGGTACACAGGTGAAGAAAGATGAAACAGTGTGCTCCAGATGAA 1022
QY 817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgThrLeu 836
Db 1023 GAGGTTTCCCTGAAAGTACGCAAGGTACAGGCAATTAACCTCTCGTAAGTGTGCTG 1082
QY 837 LeuGlyMetLysAsnHisSerLysSerGlyThrSerThrLeuArgLeuThrThr 856
Db 1083 TTGGGATGAAAAACAACAGCTTAATCTGCCAATTCACACCTTCGGTATTTATCAGCG 1142
QY 857 IleLeuHisSerArgLysAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
Db 1143 ATGTGTGTTAGTAGGGGTGACCTGCAGACAGCAAAAGGATCGAATAATCTGATATCTCT 1202
QY 877 ArgLeuArgPheAlaAlaGlySerAlaIleValLysLeuAlaGlnLysProCysThrHis 896
Db 1203 CGCTTGCGATTACCTGCTGTAGTCCATATGCAAGCTTCTCAGCAACCTTGTACAT 1262
QY 897 GluIleIleThrLeuGlnGlnThrGlnLeuCysAlaLeuAlaIleAsnAspGlyCysThr 916
Db 1263 GAATATATATCCCAACAGTTTCAAGCTCTGCACACTGTATTAATGATGAGTGTAC 1322
QY 917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
Db 1323 CAAGTAAGCGATATTTGCTCAGAAAGCTCATAAAGCAGCTGTGAAGTTACTGCTCCA 1382

QY 937 LeuGlyThrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
Db 1383 TTGGAGTATATGGCGATCTTGGCTTGTGTGGCCAAAGATCCTGTGAAGAGAGAAAGCA 1442
QY 957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgIuThrLeuLysGlnHis 976
Db 1443 CACGACGACCAATGTTTACTGAAATAATATCAGATATCGCGAGGAAATACATTAAGCAGAT 1502
QY 977 AlaAlaValSerLysLeuLysLeuLeuSerLeuLeuProGluThrValValProGlyThrIle 996
Db 1503 CTTAGGCTTACAGAAATTTATATCATCTGTTCCTGGAATATGTACTCCATCATGAT 1562
QY 997 HisLeuLeuAlaHisAspProAspThrValLysValGlnAspIleGluLeuLeuLysAsp 1016
Db 1563 CACCTGCTAGCCCATGATCCAGATTTTACAGATTCACAAGATGTGATGACGTCTGTGAT 1622
QY 1017 ValLysGluCysLeuThrPheValIleGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
Db 1623 ATCAAAAGAGCCTATGCTTCAATGTTGAAGTTTATATGCAAAAGATGAAACAATATGC 1682
QY 1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
Db 1683 CATGCCCTTATGAGAGATGGCAGAGAACATAGTTTACCAAGATGCCCAGTCTCCA 1742
QY 1057 AspAspAlaLysMetAsnGluLysLeuYThrValCysAspValAlaMetAsnIleIle 1076
Db 1743 GATGAATCCACAGCAATGAAATGAAACGTATACAGTATGATGTGCTGTCTGTATTA 1802
QY 1077 MetSerLysSerThrThrTySerLeuGluSerProLysAspProValLeuProAlaArg 1096
Db 1803 AATAGTAAAGTCTTGTGTCATGCAATGCAATGCCAAAGAGCCAGTCTCCCAAGAA 1862
QY 1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAspThrLeuProProGluMet 1116
Db 1863 TTTTTCACACACTGAAAGAGCTTCTGTAAAGATAGAGTATATTTCCAGAGAGACA 1922
QY 1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaLysAnLys 1136
Db 1923 AGAGTACTTGTGTTACAGAGAAAGCCAAAGCCTGTGAGTACTAGTGCAGTAAATAG 1982
QY 1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
Db 1983 CCTTATACCAACAGGAAAGAAACCTATGTTAGAGCACTGCGACCTGAGCTGGAAGC 2042
QY 1157 AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg 1173
Db 2043 AATATTATATTAATTCAGAGCTGAAACCTTTCAAACCGGAATGATCAAGGAAACAGAT 2102
QY 1174 LeuAspSerSerGluMetAsnHisSerGluAsnGluAsp 1191
Db 2103 TCAGAGGACAGCAAACTGTGAGTGAAGAAATGAAGAAACCTGTGAGAGATTATTTCA 2162
QY 1192 ProLeuProGlyLysLysSerAspLysArgAspSerSerLeuValAlaGlySerGluLeu 1211
Db 2163 GTCACACCTGTAAAGATATTTGAC 2186
QY 1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGluLysLeuGlyMet 1231
Db 2187 2187 2213
QY 1232 AspAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLys 1251
Db 2214 GATCAGGCTAC 2258
QY 1252 ArgGlyHisThrAlaSerGluSerAspGlu 1270
Db 2259 AGAAAGTAAACACAGCTGTGCGAGAAATATCCAAAAGAAACAGATGAGAA 2312
QY 1271 LysGluAspIleLeuGluAsnGluAspGluGlnAsnSerPro 1289
Db 2313 2313

Oy	241	LeuMetLeuGlyTyrThrSerTLeSerAspLeuSerGlnHisValPheAspLeuLeu	260
Db	791	CTGATGCTGGGAAACACATCTATACGCGATTGTCAGAGCATGTGCTTTGAATTAATTTG	850
Oy	261	GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys	280
Db	851	GAGCTACACATATATGATAGTCATTGCTGCTCTGTGTTTACCSCAGCTTGAAATTAAA	910
Oy	281	LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe	300
Db	911	TTAAAGAGCAATGATTAATGAGAGACGGCTACAACTTTTAACTACACGCGCAAAATGTTT	970
Oy	301	GluValLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly	320
Db	971	GGGCGAAAGATTAGATTGGCTTCTTCACAAACAGCCACTTTGGCAGTGGTACTTGGGC	1030
Oy	321	ArgPheAsnAspIleHisValProIleArgLeuGlnCysValLysPheAlaSerHisCys	340
Db	1031	AGGTTTATATGATATTCATCATGTCACATCCGCTGGAAATGTGTGAATTTGCTAGSCATTGT	1090
Oy	341	LeuMetLsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis	360
Db	1091	CTCATGAAACCCCTCGATTGTAGCAAAAGACTTTACAGAGATCTTAAAGTGAAGTCAAT	1150
Oy	361	AspProGlnGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys	380
Db	1151	GACCTTGAGGAAGTATTATGACATGATGTATTGTGTCAATAGTTTACAGCTGCTTAAAG	1210
Oy	381	AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys	400
Db	1211	GATTTCTCTCTGGTCATCATGATCACTTACTTAAATTTTGTGAGAGAGACACTTGACAAA	1270
Oy	401	ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla	420
Db	1271	CGAAGGAGAGACACCAAGAACCCATGATGGAGACTGCCCAATTTTAAAGAAATATGCT	1330
Oy	421	LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu	440
Db	1331	TTACAGTCACGAGCTGGAAAGATGCTGCACAAACAGATAGCATGATCAAGACAAATTG	1390
Oy	441	LeuHisIleTyrTyrGlnAsnSerTleAspAspArgLeuLeuValGluArgIlePheAla	460
Db	1391	CTACATATATATTTTTCACAAATAGATATGATGATGACTACTGTGTTGAACGATCTTTGCT	1450
Oy	461	GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu	480
Db	1451	CAATACATATGTTCTCTCCACAAATTTAGAAACTCACAAACGATGAATCTGTATATTACTTG	1510
Oy	481	TyrIleAlaPheLeuAspLeuAsnAlaValLysAlaLeuAsnGlnMetTrpLysCys	498
Db	1511	TATCCACACATCGATTTAAATAGCTGTAGATATTTTCAATATCAATATTCCTGT	1564
RESULT 10			
AK098331			
LOCUS			
DEFINITION	AK098331	2265 bp mRNA linear PRI 15-JUL-2002	
		Homo sapiens cDNA FLJ41012 f1s, clone U7802018153, highly similar	
		to Human androgen-induced prostate proliferative shutoff associated	
		protein (AS3) mRNA.	
ACCESSION	AK098331		
VERSION	AK098331.1	GI:21758323	
KEYWORDS		Oligo capping, f1s (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE			
AUTHORS	1	Kawamali,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,	
		Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,	
		Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,	
		Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,	
		Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,	
		Oshima,A., Suzuki,T., Sugano,S., Negaharui,K., Masuno,Y., Negai,K.	
		and Isogai,T.	

TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2265)
AUTHORS	Isogai,T. and Yamamoto,J.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, Fij Project(HRI Team); 2-6-7
COMMENT	Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:igomios@hri.co.jp, Tel:81-438-92-3975, Fax:81-438-52-3966 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB. Location/Qualifiers 1. 2265
FEATURES	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="UTERU2018523" /tissue_type="uterus" /clone_lib="UTERU2" /note="cloning vector: PME18SFL3" 125..1927 /note="unnamed protein product" /codon_start=1 /protein_id="BAC05286.1" /db_xref="GI:21758324"
SOURCE	/translation="MDFTAPKPTALCGVVSADGKTAIPPGVKETDKTTDEMIRK LKNVATFEMDQDSEDERQOYLPLALHAEFLRNPNKNDVRLVACCLADIFRITY PEAPYSHDKLDFLFTTRKGLGEDRSPQPNRYLLENLAWSVKINCELED NEEFLDFEFLTELVINSINKVOMHMLDSSIMEGDVTOLDSIILNIPAH NKKOSFEDLAKVLKRVQTEITACIANTFANVQLVIGSSVDSLEHYEDLOELFPAH PHLLSVMPLEFLKLVNDGERRLAVVRLAKLFGNSDOLATONRPMDFCGFRFN ALYHVPLESEFKASHCNHNPDLADLLEYIKAVSHDEPARIDVYVITTAARR ALYNDOLGFRERFLDKRMVRKAMMGLKQTKTCLHGEAKREAEVSWTKD LHYVONSIDDKLVEKLFPAOYLPHNLETEERKCLYLYIASIDPAVVALMMKK KNMRSHVRELLDLHKOPTSEANCSAMFGKLMTITAKNLPDPGKADQVKKENYVLG EKLRSOLELLISPTCSKQADICVCKSYFTLFL"
CDS	BASE COUNT 696 a 427 c 476 g 666 t ORIGIN
Alignment Scores:	
Pred. No.:	2.94e-122 Length: 2265
Percent:	2474.00 Matches: 460
Score:	91.77% Conservative: 64
Best Local Similarity:	80.568 Mismatches: 47
Query Match:	34.39% Indels: 0
DB:	Gaps: 0
US-09-512-581B-2 (1-1391) x AK098331 (1-2265)	
QY	10 AspGlyLysIleThrTyProProGluValLysGluIleSerAspLysIleSerLysGlu 29
Db	182 GACGGAGAGATCGCTTACCTCCGGGGGTAAGAGATACCGCAAGATACACGAC 241
QY	30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSer 49
Db	242 GAGATGATCATCAAGCGCTGAAGATGGTAGTGAAGAAACCTTTAAGTATGAGTACAGACTCA 301
QY	50 GluGluGluLysGluLeuTyLeuLysLeuAlaLeuHisLeuAlaSerAspPheLeu 69
Db	302 GAAAGATGAAAACAGACAGATATCTCCCACTAGCCTTGCGATCTTGATCTTCTTCTC 361
QY	70 LysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
Db	362 AGGAACCCCAATTAAGATGTGGCGTCCTTGAGCATGTGTTGGCGATATCTTGCT 421
QY	90 IleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPhe 109
Db	422 ATTATGACCCCAAGAACTTCATATCTTCCCATTAAGTAAACATATATTTGGTTT 481

QY	110	IIETHTAAGGInLeuLysGlyLeuGlnAspThrLysSerProGlnPheAsnArgThrPhe	129
Db	482	ATTACCAACAACAAATTTAAAGGTTTGGAGGATACAAAGAGCTCCACAGCTTTAATAGATACCTT	541
QY	130	TYTLeuLeuGlnAsnILEALATPrALysSerTYrAsnILECysPheGluLeuGlnAsp	149
Db	542	TATTATTATAGAGAATTAGCTTGGGTTTAATCATATTAACATCTGCTTGAATTGGAAAT	601
QY	150	SerAsnGluILEPheThrGlnLeuTYrArgThrLeuPheSerValILEAsnAsnGluHis	169
Db	602	TGCATAGAAATTTTATATTCACCTTTTGTGAAGACTCTCTTCAGATGATCAACAATAGCCAC	661
QY	170	AsnGlnLysValHisMetHisMetValAspLeuMetSerSerILEILECysGlnGluLysp	189
Db	662	AATAAGAAAGGTACAAATGACATGCTTAATTTGATGAGACTTCTATCATATGAGAACTGAT	721
QY	190	ThrValSerGlnGluLeuLeuAspThrValILEuValAsnLeuValProAlaHisLysAsn	209
Db	722	GGAGTTACTCAAGAAATTATTTGGACTCAATCTCTATTAACTTCATCTCCGACATAGAAC	781
QY	210	LeuAsnLysGlnAlaTYrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaILE	229
Db	782	TTAAATTAACAGTCCCTTTGACCTTGCAAAAGTGCATTTGAAAGAACAGTCCAGACTT	841
QY	230	GluProTYrILEThrThrPhePheAsnGlnValLeuMetILEGlyLysThrSerILESer	249
Db	842	GAGCAGTCATGCTGCTAATTTTTCATGATAGACTCCGTGGTCTGGAGAGATCATCAGTAAT	901
QY	250	AspLeuSerGlnHisValPheAspLeuILEuGlnLeuTYrAsnILEAspSerHisLeu	269
Db	902	GATTTGTTCAGAAACATGTATTGATCTGATTCAGAGAACTTTTGTGATAGATCCATCTTA	961
QY	270	LeuLeuSerValILEuProGlnLeuGlnPheLysLeuLysSerAspAsnGlnGluArg	289
Db	962	TTATTATCCGATACCCACACAGCTGTGAATTCAAACTAAAGACAAATGATGGAGAAAGCA	1021
QY	290	LeuGlnValValLysLeuLeuAlaLysMetPheGluAlaLysAspSerGluLeuLaser	309
Db	1022	TTAGCTGTGTTGCGACTCTGAGCTAAATTTGTTGGCTCCAAAGATTTGATTTGGCAACA	1081
QY	310	GlnAsnLysProLeuThrGlnCysTYrLeuGlnArgPheAsnAspILEHisValProILE	329
Db	1082	CAGATCTCCTCTTTGGCAATGTTTCTTGACACATTTAATGATATTCATGTTCTCTTG	1141
QY	330	ArgLeuGlnCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys	349
Db	1142	AGATTAGAAAGTGGAATTTGCCAGCTATTTGTTAATGANTCACCAGATTTAGCGAAG	1201
QY	350	AspLeuThrGluTYrLeuLysValArgSerHisAspProGluGluAlaILEArgHisAsp	369
Db	1202	GATCTCAAGAGATTTTAAAGGTAGATCAACATGTATCCAGAAAGAGCTATTGCTCATAT	1261
QY	370	ValILEValSerILEValThrAlaLysLysAspILELeuLeuValAsnAspHisLeu	389
Db	1262	GTCATTGTACTATATATACAGCTGCCAAGAGGAGACCTGGCTTAGTAAATGATCAGCTG	1321
QY	390	LeuAsnPheValArgGluArgThrLeuAspLysArgTYrParValArgLysGluAlaMet	409
Db	1322	CTTGCGCTTTTGAAAGGAAAGAACCTGATTAAGCGTGCCAGATTAAGAAAGAGCTATG	1381
QY	410	MetGluLeuAlaGlnILETYrLysLysTYrAlaLeuGlnSerAlaAlaGlyLysAspAla	429
Db	1382	ATGGGTCTGGCTCAGCTTATATAGAAATACGTCTTCACTGGAAGCAGGAAGGAGACT	1441
QY	430	AlaLysGlnILEATPrILELysAspLysLeuLeuHisILETYrTYrGlnAsnSerILE	449
Db	1442	GCAAGAGAAAGACACTGGATTAAGAGACAAACTCTGCTATTTATTATCAGAAACGACTT	1501
QY	450	AspAspArgLeuLeuValGluArgILEPheAlaGlnTYrMetValProHisAsnLeuGlu	469
Db	1502	GAGCAACAACGTGTGTAGAGAAAAATCTTCTCAGTATCTTGTGCCCCCAGAACCTGGAA	1561

QY	470	ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaIleAlaMetLeuAspLeuAlaVal	485
Db	1562	ACAGAAAGAGAAATCAAAATGCTTAATTACTTATATGCTACTCTTGCAATCCAAATGCTGTA	16211
QY	490	LysAlaLeuAaGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu	509
Db	1692	AAAGTCCTCAACGAATATGTGGAAGTGTGAGAACATGCTTCGGAGCAGCATGACGGAACTA	16811
QY	510	LeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal	529
Db	1682	TTGGATTTTGGCAACAGCACCTTACATCAAGAGCTTAACCTTCCATCTTTGGAAAACTG	17411
QY	530	MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe	549
Db	1742	ATGACCAATGACAAAGAAATTTGGCTGCTGACCCCGGAAAGACACAAAGATTTTGTACAGAAATTT	18011
QY	550	ThrGlnValLeuGluAspGluLysIleArgLysGlnLeuGluValLeuValSerPro	569
Db	1802	AACCAAGGTTTCGCGGATGATGACAGAAACTTGGCTGCTGAGATTTATTAATAGCCCA	18611
QY	570	ThrCysSerCysLysGlnAlaGluLysVal	580
Db	1862	ACCTGCTCTTGCACAAACAGACAGATATTTGCTGTG	1894
RESULT 11			
LOCUS	BC041361	2705 bp	mRNA
DEFINITION	Homo sapiens, similar to KIA00648 protein, clone MGC:43818		
ACCESSION	BC041361		
VERSION	BC041361.1		
KEYWORDS	GI:27552793		
SOURCE	MGC.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	I (bases 1 to 2705)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
REMARK	Submitted (16-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: http://mgc.ncl.nih.gov		
	Contact: MGC help desk		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
	cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLN)		
	DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada		
	info@gscc.bc.ca		
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guh, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandohs, Anna-Liisa Prabhun, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miraneda Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLN at: http://image.llnl.gov		
source	Series: IRAK Plate: 75 Row: c Column: 19.		
	Location/Qualifiers		
	1..2705		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="MGC:43818 IMAGE:5273075"		
	/tissue_type="Testis"		
	/clone_id="NIH.MGC.97"		

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT		
BC021408	Mus musculus RIKEN cDNA 9030416H16 gene, mRNA (cDNA clone MGC:29439	BC021408	1	GI:18204099	Mus musculus (house mouse)	Enkayrta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	1 (bases 1 to 1976)	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Canninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McEran, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Moley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylar, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Fahey, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	22388257	12477932	2 (bases 1 to 1976)	Strausberg, R.	Direct Submission	Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland. Web site: http://www.nisc.nih.gov/ Contact: nisc.mc@nih.gov Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McQuill, J., Pearson, R., Stantirlop, S., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L. H. and Green, E.D.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Series: IRAC Plate: 39 Row: B Column: 24 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.	Location/Qualifiers 1. 1976 /organism="Mus musculus"

[illegible]

Db 445 GAAGGAGCGCCGATGATGAAGAGGAGGTGTCTAGTCCAGATCCGATCCGCTCAGGA 504
 QY LeuGluleuLeuValleuSerPheThrHisProIleSerPheHisSerAlaGluThr 669
 Db 505 CTGGAGCTTCTTAAGGTCTGTCTTTCACACATCTCACTGCTGCTGCAGAGACA 564
 QY PheGluSerLeuLeuAlaCysLeuLysMetAspSerPheValAlaGluAlaLeu 689
 Db 565 TAGAGAGCTGTGATACGCTCCTAAGAAAGGAGATGACAAAGTAGAGAGAACTGCATATA 624
 QY GlnIlePheLysAsnThrGlySerLysIleGluIleAspPheProHisIleArgSerAla 709
 Db 625 CAATATTTTGAAGAACACAGGCCACAAAATAGAAACCTCCCTCCAGATACGGTCAACC 684
 QY LeuLeuProValleuHisIleLysSerLysLysGlyProProArgGlnAlaLysTyrAla 729
 Db 685 TTGATCCCATTTTACTTCAGAAAGCCAGCGGGGACCTCCACACCAAGCAAGACAGGCT 744
 QY IleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGlu 749
 Db 745 GTTCACATGCATGCATGCTCTCTCAACAGAGAGGTCCAGCTGCAGATTTTGTAG 804
 QY ProLeuHisLysSerLeuAspProSerAsnLeuGlnIleuIleThrProLeuValThr 769
 Db 805 CCACATGATGAGAGTCTGAATGCTGATGACACAGCAACTTATTAATCTCATTAAGTTTCA 864
 QY IleGlyHisIleAlaLeuAlaProAspGlnPheAlaIleProTyrPheLysSerTyrVal 789
 Db 865 CTGGGCCCATTTCCATGTTAGCAGCCAGATCAGTTTCCCTCCCGATGAATCTGTAGTG 924
 QY AlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThr 809
 Db 925 GCAAACTTTATTTGTTAAAGACCTTCTAATGAACGACAGCTCAACAGCTGAGAAAGATGGA 984
 QY LysLeuTyrValProAspGlnGluValSerProGlnIleuMetAlaLysIleGlnAlaIle 829
 Db 985 AAATATGAGTCTCAGATGAGAAAGTTTCTCTGAATCTTACGAAGGATGAGGCAATTT 1044
 QY LysMetMetValArgTyrPheLeuGlnLysMetLysAsnAsnHisSerLysSerGlyThrSer 849
 Db 1045 AAATCTCTGTAGTGGCTGTGGGTGATGAAAACACACAGCTTAATCTGCCCACTCA 1104
 QY ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlnLysPheLeuThrGlnGlnLys 869
 Db 1105 ACCTCTCGTATTATATACGCCATGCTGGTTAGTGAGGCGCAGCCGACGAGCAAAAGAG 1164
 QY IleSerLysProAspMetSerArgLeuArgLeuAlaIleAlaGlySerAlaIleValLysLeu 889
 Db 1165 ATCAGTAAATCTGATATGCTCGCTGGCATGTAGCTGCTGTAGTCCATATGCAAGCTT 1224
 QY AlaGlnGluProCysTyrHisGluIleIleThrLeuGlnGluIleuThrGlnLysCysAlaLeu 909
 Db 1225 GCTCAGGAACCTTGCTACATGAAATATTACCCAGAACAGATTTCAGCTGTGCACTG 1284
 QY AlaIleAsnAspLysCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGly 929
 Db 1285 GTTATTAAATGATAGGCTACCAAGTAAGCAGATATTTCGCCAAGAGCTTCTAAGCA 1344
 QY LeuSerArgPheArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 949
 Db 1345 CTGTGTAAGTGTCTCTCCACTGAGATATATGCGCATCTTGTGTGTGTGTCNAAGAC 1404
 QY ProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAspValArg 969
 Db 1405 CCTGTGAAGAGAAAGAGAGACACGCTCGCAGAGTGTGTTAAAGAACATCAGATCCGC 1464
 QY ArgGlnTyrLeuLysGlnHisAlaAlaValSerGlnLysLeuSerLeuSerLeuProGlu 989
 Db 1465 AGGAGATACATCAAAACAGAACCCCAAGCCAGAGAGAAATATTGTACAGCTGCCCGAA 1524
 QY TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGln 1009
 Db 1525 TATGTGTTCCATACATGATTCACCTCTAGCCCATGATGCTGATTTGCACACACACAA 1584

QY 1010 AspIleGluGlnLeuLysAspValLysGluCysLeuThrPheValleuGluIleLeuMet 1029
 Db 1585 GATTTATCACTTCTGATATTAAGAGAGTCCGTGGTTTATGCTTGAAGTCTTATAG 1644
 QY 1030 AlAlasngLusnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGln 1049
 Db 1645 ACAAGAATGAAACACACAGCATGATTCATGAAAAGATGGCAGAGAAATATCAACTA 1704
 QY 1050 ThrLysAspAlaGlnGlyProAspAspAlaLysMetLsnGlnLysLeuTyrThrValCys 1069
 Db 1705 ACCGAGACGCCCGCTCCCGATGATCAAGCAAAATGAAAACCTTATACGTTTGT 1764
 QY 1070 AspValAlaMetAsnIleIleMetSerLysSerThrThrLysSerLeuGlnSerProLys 1089
 Db 1765 GACCTGCTCTGTGTGTTATTAATAGTAAAGTCTTTGTGCAATGCAGACTCACCAAG 1824
 QY 1090 AspProValleuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThr 1108
 Db 1825 GACCCAGTCTCCCAATGAAGTCTTATACACAGCTGAAAAGGTAATTCCTCCACA 1881
 RESULT 13
 AL844559
 LOCUS 216608 bp DNA linear VRT 28-JAN-2003
 DEFINITION zebrafish DNA sequence from clone DKEX-14K1, complete sequence.
 ACCESSION AL844559
 VERSION AL844559.14 GI:28144648
 KEYWORDS HMG
 SOURCE Danio rerio (zebrafish)
 ORGANISM
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE
 1 (bases 1 to 216608)
 AUTHORS Babbage, A.
 TITLE Direct Submission
 JOURNAL Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
 On Jan 30, 2003 this sequence version replaced gi:27801645.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C.elegans/wormpep/ Repeat names
 beginning 'Dr' were identified by the Recon repeat discovery system
 (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dr'
 were identified by Rick Waterman (Stephen Johnson lab, WashU). For
 further information see http://www/projects/D_rerio/fishmask.shtml
 DKEX-14K1 is from a zebrafish BAC library
 VECTOR: pIndigoBAC-5.
 FEATURES
 Location/Qualifiers
 I. .216608

```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-14K1"
/clone_lib="Daniokey"
BASE COUNT 68345 a 39077 c 40063 g 69123 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 6,79e-86 Length: 216608
Score: 1836.00 Matches: 916
Percent Similarity: 18.40% Conservative: 148
Best Local Similarity: 15.84% Mismatches: 215
Query Match: 25.52% Indels: 4511
DB: 5 Gaps: 39

```

US-09-512-581b-2 (1-1391) x AL844559 (1-216608)

```

OY 37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56
Db 95986 TTGGGGTTAAGACATATATGACATGATGATCAAGACTCAGAAAGAGAAAGACAGATAC 96045
OY 57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPheLeuLysHisProGlyLysAspVal 76
Db 96046 TTGGCTTTGGCTTCCACCTGGCTCTGAGTCTTTCTGCGAACCACCAAGGATGTA 96105
OY 77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
Db 96106 CGGTACTGCTGGCTGGCTGCTGCTGCTGACATCTTCAGATCATATGATCCAGAGCACCA 96185
OY 97 TyrThrSerProAspLysLeuLys----- 104
Db 96166 TACACCTCAGATGATAAAGCTCAAGGTATACCAATACATCATCTAGCTGCTTGAGAA 96225
OY 104 ----- 104
Db 96226 ATGTGTAATCTTTCAAAAGACTTTTACGTTGTCMAACTGATTTCTTCTACCTGCAC 96285
OY 104 ----- 104
Db 96286 ACAATCACCTCAGCTATACGAAACCTTAATGTGATATTTTCAGATGTTAAAGACAGTAC 96345
OY 105 -----AspIlePheMetPheIleThrArgGln 113
Db 96346 AGTGATGATGGTATTTGTTTTCCTATCTAGAAATTTTCTTTTCATCAGAGGACAG 96405
OY 114 LeuLysGluLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuGlu 133
Db 96406 TTGAAGGGCTTTGAGGACACCAAGAGCCACAGTTCAACAGATATTTTACTTGTAGAG 96465
OY 133 ----- 133
Db 96466 GTGAGATGGCTTTTAATGAGAGTATTATGATCTTTTGTCTCTGTTTACAGAAAGAA 96525
OY 133 ----- 133
Db 96526 AGGCTTGAATGCTTTTAAATTACTCTCAGACGCAAGGCATCTAGAGCTTTTGTG 96585
OY 133 ----- 133
Db 96586 ACCAGTTTGGCTTTGTTTATGATCGTATTATGATTTGTCAGAGCTAACTTTTCAT 96645
OY 134 -----AsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAs 149
Db 96646 TTGACTTGACAGAAAGCTTGGCATGGGTGAAGTCTCTACAAACATCTGTTTGAATTGAGGA 96705
OY 149 pSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn----- 166
Db 96706 CTGTAAATGAATCTTTATCCAGCTCTTCAAAACCTCTTTCTGTGATCAAA-GTGAGTTA 96764
OY 166 ----- 166
Db 96765 ATTTGAGACGATGTAATTATCCCAATTTGTTCTCAAGCTGTCTTTCTGATCAGAC 96824

```

```

OY 167 -----AsnGlyHisAsnGluLysValHisMetHisMetValAspLeuMet 181
Db 96825 AATGCACCTTTTCAATTAACAGGCATATACAGAGGTGCAATGACATGTTGTGACTGAT 96884
OY 181 tSerSerIleIleCysGluGluAspThrValSerGlnGluLeuAspThrValLeuVal 201
Db 96885 GAGTTCATCATCATGAGAGGTGATGATGATGACGCAAGAGCTGTGACACAAATCCCAT 96944
OY 201 tAsnLeuValProAlaHisLys----- 208
Db 96945 CAACCTCATCCCTGCTCACAAAGTAGGCAAAATAGACCTATTGTGATTTATTTACATTCCA 97004
OY 209 -----AsnLeuAsnLys 212
Db 97005 GTTCAGATAGCTAACCAATATAGGTTTCAACCATTTCTGTGCTTTAGAACCTTGATTA 97064
OY 212 sGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyr 232
Db 97065 GCAGGCATATGACCTGGCCAGGACTCTTCTTAAGAGAAAGGTACAGACTATCGAAACCTG 97124
OY 232 tIleThrThr----- 235
Db 97125 CATCCCTCAGTGAAGTACAAAAATCATCTGTTTGTAAATTATTTAGATAATAGTTT 97184
OY 235 ----- 235
Db 97185 CAGCATTTAGCATGCTCAATATTATTTCATATGTTGATTTGCTTGGCTTGGCATTTCTTTCT 97244
OY 236 -----PhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHis 254
Db 97245 GTAGTTCTTCATACAGGTGTTGCTGATGGAAGATCCCGGGAGTGAATCTGTACAGACA 97304
OY 254 sValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLe 274
Db 97305 tGTATTGACCTCATTCAGGAAGCTCTTTCATATGACCTCTGCTCAGTCTGTGAT 97364
OY 274 uProGlnLeuGluPheLysLeuLys----- 282
Db 97365 GCCACAGCTTGAGTTTAATTGAAGTGACCAAAATTGACTCAGACTTAATTTCACACCAG 97424
OY 283 -----SerAs 284
Db 97425 CTTCATACAGAGCTTTTAGAGTAAGTCTTTTGTGTTTGTCTTTGCTTGTACAGTAA 97484
OY 284 nAspAsnGluGluAlaGluGlnValValLysLeuAlaLysMetPheGlyAlaLysAs 304
Db 97485 TGATGGAGAGGAAAGTTAGCTGTGTAATAATTCTACCAAAACTTTGGCGCAAAAGA 97544
OY 304 pSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly----- 320
Db 97545 CTCTGAACCTGGCACTCAGAAATGCCCACTCTGGCAGTCTTCTGGGAGGTAGGATGG 97604
OY 320 ----- 320
Db 97605 CTTTACCATTTAAATCTGTAATTAAAGTGTAGAGATTTGAGACTTTCTTAATGAA 97664
OY 321 -----ArgPheAsnAspIleHisValProIleArg 330
Db 97665 TTTGGCTTTAATTTTGTCTTTTGTCTTGAAGGTTCATAGACATCAATGTTCCGTGAG 97724
OY 330 gLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
Db 97725 ACTGGAATGTGAAAGTTTGGCAGACACTGCTGATGAACCATCAGACCTTGGCAAGGA 97784
OY 350 pLeu----- 351
Db 97785 TCTTAACAGTCTGTTCTGCTGAATACAAATGCTGTTGTTTCTATTTTAAGCACT 97844
OY 352 -----ThrGluTyrLe 355
Db 97845 CTTGACATAGAGATGCTGTTTAACTTAATGATGTGCTATATTTCACATACAGAGTTT 97904

```

D	b	98962	ATACTATATCTTAGTCTTATGTTGGTTTAAATAAGAAAAAATATTTTGTGCTTAAT	99041
Q	y	516	-----	516
D	b	99042	ATAGTTTATTCAAATATGTAGTTACTGATATATATATATATATATATATATATATATA	99101
Q	y	516	-----	516
D	b	99102	TA	99161
Q	y	516	-----	516
D	b	99162	TGTGTGTATGTATGTATTTGCAGACATTAGTCTTGCAGGTATGTATGATCCCTCAGTG	99221
Q	y	516	-----	516
D	b	99222	TGATGAGCCTTGCTGCTGCTCTGATGCGATGATCAGCTGCGTGTATTAACATTTCAC	99281
Q	y	516	-----	516
D	b	99282	AACAGAAAAAGCTCAATGTGAATTATTTCTGACGTTTCAGTATCTGACAAAGATCGACAG	99341
Q	y	516	-----	516
D	b	99342	TAGCAGACAGCTCGACTTAATGTGCTGTATGTATCATTAAATTGAGCCGCTTAATGT	99401
Q	y	516	-----	516
D	b	99402	TCCGTGTAACTCTGATTTAAAGCTTTAAAGCATACATCATGATGTTCTGCACCTTA	99461
Q	y	516	-----	516
D	b	99462	AAAAAGACTGATTTGGTACAGTGACACTATACGAAAGTCTGTACTGACACAGTTTAC	99521
Q	y	516	-----	516
D	b	99522	AAATATACCCAGACACTTTTACTTCTCTGTGTGTTTACATGTCTAAAGCATTAATACG	99581
Q	y	516	-----	516
D	b	99582	AATGACATACCTAGACAGCAAAATATCTGCAAGTATGAGACTATACATACTAATTAGT	99641
Q	y	516	-----	516
D	b	99642	AGTGAATAATGTTGTGATGTTCTATTTTGTAAAGTGTACATAGACAGAGTTTCC	99701
Q	y	517	-----	517
D	b	99702	ATATATGATGACTGACTTTGTTTTTCCCTTTCCGAGAGTGTGAAGCAACACTCAG	99761
Q	y	524	lailepheserlyslvalmetvalile-----	532
D	b	99762	CCATGTTCCGGAACCTTATGACCATCGCTAGTAAGCAATTCATATATTCTGTAAATT	99821
Q	y	533	-----	533
D	b	99822	GATAATCGACAGTTTAAACATGATTAAATTGGCTCTCTCATCTCAGAGAAATCTAC	99881
Q	y	537	roaspprogllylsalaglnasphnemellyslsyshethrghlnvallleugluasasp	557
D	b	99882	CAGACCCAGGCAAACTCAGGACTTATGAAAGATTATCAGAGTTCTGGGTGAAGACG	99941
Q	y	557	luylsillelryglslneugluvalleuvalserprothrcyssercylsyglalag	577
D	b	99942	AGAAACTAGTCTTACGTTGAGCAACTCATAGCCCACTGCTCATGTAAACAGGCCG	100001
Q	y	577	luqllycysval-----	580
D	b	100002	AACAGTGTGTGAAGCTTAAAGAGCCCATATTATGGTTTTTGAATAATTCCCTCCATG	100061
Q	y	580	-----	580

Db	100062	TAGTGTGTAACACAGCCTTAAGTGAAGTAAGTATTCACGCTAAAGCTTAAATCTGTAA	100121
QY	580	-----	580
Db	100122	ATACAGTGTTTAAAACTGTGTGATTCATCATTAAGAGTCGATCATAGTGGTTCAACG	100181
QY	580	-----	580
Db	100182	AGTGGCCTTGATACCGACTCATTAAGTGTTCGCCATGACGTACGAAACGAAAGTTT	100241
QY	580	-----	580
Db	100242	TTCACTGTGCACGCCCAAAACCGGGAGATTTCAAACCTGAGGCCCGCCTGTACGCACT	100301
QY	580	-----	580
Db	100302	AACCCAGACACACACACACACACACCACAACACACACACACACCCACAACACACGACACA	100361
QY	580	-----	580
Db	100362	CAACATGCGCGTCGATTGAAGTACACACTGCAGATGATATATTTGAGTCTCATCCAAAG	100421
QY	580	-----	580
Db	100422	ATGAAACTCAGACATTATTAACCAACAGCAGTTGAAAATCTCTGGAAGCTACACTACCAAG	100481
QY	580	-----	580
Db	100482	AATACTTCACTGAGTTTGTTAAGGAAGATCACTAAAGACTAATGATGACGTCAGG	100541
QY	580	-----	580
Db	100542	ATGGGTTTCTCCATACATTCTCAAGTGAAGTACGAGCGGTTAAAGTTTGCCTCGTT	100601
QY	580	-----	580
Db	100602	TACTCTAGCCTTCAAAATGATTAATTAAGTGTGATTTGTACTGTAAACCGGTACTGTAT	100661
QY	580	-----	580
Db	100662	CAGGTAACTGGATTAATTGTGATCTTATCCGCTGCAAAAGTCAGCTTAAAAACGGAGCGGTG	100721
QY	580	-----	580
Db	100722	CTGTTTGTATTGAGAGTCGCGTGTAGAGTTCCTGCCGCGATTATTCGGAATGTAT	100781
QY	580	-----	580
Db	100782	TCCCGCGCCGAGAAATCTGGCGCGGGAGAGAGAGAGAGAGAGCGGTAGGAG	100841
QY	580	-----	580
Db	100842	AGCGAAGCAAGCAAGACACTTTGTGTACTTGTGTCTGTGTGTGTGTTTATACTGACAG	100901
QY	580	-----	580
Db	100902	CTTGCGTGTGAGACTGTATACGTGGGTATATTGGTGTGTCTCCGCTACGCGGAC	100961
QY	580	-----	580
Db	100962	CGGCGCGCGCGGCGAGAAAACGAGCGGGTCTCAGGCTAAAACCTGGCGGGTGGGCGCT	101021
QY	580	-----	580
Db	101022	GAAACGGAGAGGTTGTCAATCCGGAGGTGTGTGTGTTTGTGTGTGTGTGTATGGCG	101081
QY	580	-----	580
Db	101082	AGTACACGACTGTCTCTTCGTTGCGTTATCCGTGACACTATCACTGCCCATAGTGTG	101141
QY	580	-----	580
Db	101142	CAGCTAAAGTCACGCTACACTATCGAGCGGTGTGAAGTGTGATTTTAAATTTG	101201

QY	580	-----	580
Db	101202	CTGATTAGCTTTGGCCATTTCCTCTCTGACTGAAGGCACTGCAGCAATCGCAGAC	101261
QY	580	-----	580
Db	101262	TGTCATCGGTCCAATCAGCGCAGATTAGCTTCGGCTTAAGGAGGGTTTGGGAACCAATG	101321
QY	580	-----	580
Db	101322	AATCACGTGACGATTCATACAGAGATGCGTGGATAATTAGATATAAATAATGCAGATT	101381
QY	580	-----	580
Db	101382	ATAGACCATGAAGTGTTTTTTTTGACCTTGATCATATTAGACTGTTGTTGAGACCT	101441
QY	580	-----	580
Db	101442	TACAACCAAGATATGACCCATTTCATGTATATATGSGCTTTTAACAAACTTTTTTT	101501
QY	580	-----	580
Db	101502	AATGCTGTGTCAAAACTCGGCCATGCTTCATTAACAGGGCTTCACATGACTTTTTT	101561
QY	580	-----	580
Db	101562	GATCACCGCCACACAGTGGCTATTAGTTTTCACAGTTACTAGCCACTCTGCATTTCCT	101621
QY	580	-----	580
Db	101622	AGCCACATTTTGTGTGGGAAAATATTTTATATGCACAATATTACTTGCATGC	101681
QY	580	-----	580
Db	101682	TAAATTAATCTGATTTAGATTTTGTGTTTGTACATGCCGCTCATTAATGCACGTT	101741
QY	580	-----	580
Db	101742	TTGTGTCTCGTGATGAGCAATGATGATGTTTCATAGTGTCTATTACAAATTAGTT	101801
QY	580	-----	580
Db	101802	TTATTTACAGATTTTTCAGAACTCAAAATTAGGATGTACAAATTTATCTGCACA	101861
QY	580	-----	580
Db	101862	CTCCGACAAATATATTATTCTTAGCCACACGTTTAAATGGTCAAAAACAAGTAAATAT	101921
QY	580	-----	580
Db	101922	AGCTGATATATTACTTTTATTTAACAAAAACATTTCTTAAAAAAGAAAAAAG	101981
QY	580	-----	580
Db	101982	AAAAACATTGACTTTTAAAAATAGTTATAGTCATGTATCAAAAATATGCAGATAATCT	102041
QY	580	-----	580
Db	102042	GTCCTCGTAAAGTCTCTCAGATCACTAGTTACTACACAAGTGGAGTTAATCTGATA	102101
QY	580	-----	580
Db	102102	TTAAAGCATGTGTTGTAAAAAATGATGATTAACATTAATAACAAAAATAACCGTAAG	102161
QY	580	-----	580
Db	102162	CAAAAGAAAGCAGTGAAAAACATACGTGTGATTAATGAAGAAGATGATCGATGTTTTG	102221
QY	580	-----	580
Db	102222	GCCATTAATATCTGTGTCAAAAGTAAAGCAACTGTGTGCTGTACAAAAAGACCAATCT	102281

QY 580 ----- 580
Db 102282 GGAAGCTTGAAGACAGCAGACTGCGGGTGCAATTAGCATTAATTTTGTCTGTATATA 102341
QY 580 ----- 580
Db 102342 TGAAGAGAACTGATTAACAGTTCGGTCCAGGACGATTCGTTCAAGCAAAACAAGAGA 102401
QY 580 ----- 580
Db 102402 GCGGACACGGGGTTTATAAGTCGGCTGTATCATCAGCTGTGTGCAGAGACTGATCATCT 102461
QY 580 ----- 580
Db 102462 CTCATTCATTGAGTCCAGATCATCCAGATATTACCTGCTTCTATTGTGCGCGGAAA 102521
QY 580 ----- 580
Db 102522 GCAAAATTATTTTGTAGTGTGCTGTTCTCGATTCTGAAGATCACAATTTGCACGCATAT 102581
QY 580 ----- 580
Db 102582 TGGGTATCCTTAATAGTCGAACCTGCTTTTAAGAAAACATATTTAAATTTGTTTC 102641
QY 580 ----- 580
Db 102642 AACCACTGAAGTGGCTAGTGGGTAGGCTGCACGATATATCGTTTCAGCATCGAAATCG 102701
QY 580 ----- 580
Db 102702 CGATGTCCGATCCGCGATAGTACATCGACAGAGGGTCCGATCCAAAGCAAAAAAAA 102761
QY 580 ----- 580
Db 102762 AAAAATTTCTATATATATTATGGGTGAGCGCAACCGAATATGTATTTCAGAAAAGCAG 102821
QY 580 ----- 580
Db 102822 AATATATGTTTTTACGAATACTATTTCAAACATACCTGTATTAATTTTGTATTTGGG 102881
QY 580 ----- 580
Db 102882 AACAGAAAAACATTATATCAAAAAAGCTGCGTTTCTTAATGAGACCCAGTCATGCCCG 102941
QY 580 ----- 580
Db 102942 CGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGTGGGGTGGGGCCGGGGA 103001
QY 580 ----- 580
Db 103002 AGGATCTGAGCAAGCTGCTTTCTATACAGCAACAGAAAGGCTTGGCAGAGCAAAAA 103061
QY 580 ----- 580
Db 103062 TACTTTTCTGCAATTACTATTTTGTATTATTTATCACTAATAACAGCTGACAGAAG 103121
QY 580 ----- 580
Db 103122 TTCACTCAAAAGTAAAGTTCTTCTGTATTAATTAACAACACTCCCGCATTCCTACT 103181
QY 580 ----- 580
Db 103182 ACTTTCAGAACATTAATTTTGATATTTTACCATATGAATAATGCAATGCAGTTATGAATA 103241
QY 580 ----- 580
Db 103242 TGAATAATTAAGCAATACACAGATCTGTCCCGGCTCAGGTACGGGTGCACCTGTAA 103301
QY 580 ----- 580
Db 103302 CTTAAAAAGCAGACGGGGGAGGGCGCACTAAACTCAACAAGGAAAGAACTGGGCATTC 103361
QY 580 ----- 580

Db 103362 AATTCATGTTTGTGTTAAATTAATCTTAAATACAGCTCTTTGATCTTAAGGGAAAA 103421
QY 580 ----- 580
Db 103422 AAAACGAAAGAAATTTGTGCTTAATTAATTTTACGAAAGCAGCACTGCTGCATTT 103481
QY 580 ----- 580
Db 103482 TTCAATTGTTTTTCAGTTAGGGATTTGGCTATTAGAGTAACCTAATAAATTTGTTTT 103541
QY 580 ----- 580
Db 103542 GTTGGACACGTAGTCTACACAATTAATTAATTAATTAATTAATTAATTAATTAATTA 103601
QY 580 ----- 580
Db 103602 AAACTAACCTATACATTTGATTAACAAAAACCCCTTTTAATGTTTGTCTTCATTTTGACA 103661
QY 580 ----- 580
Db 103662 CCGCTACAGTGTGAATAGAAGAACATCATAATTAAGCAGATGCAAGTCAAGTTAAA 103721
QY 580 ----- 580
Db 103722 AACACCTTTGAGTTTAAGATGATCAGCTCGAAGTCAAGTTTACAGTGTATTAATG 103781
QY 580 ----- 580
Db 103782 AATGGGTACATGATGAAGGGGATTTGCCACAGCAACAACCTCATGCGCTGACGGTAA 103841
QY 580 ----- 580
Db 103842 GCTGAATGATTAATCTTAATAAATTAATAGCTATGCAATACAAACTTTAAACATTTAT 103901
QY 580 ----- 580
Db 103902 TATTTTTTAATGAAGAAAAAGAAACAGAGAGTAATAATTAATTAACACACTGCTCT 103961
QY 580 ----- 580
Db 103962 GGGCTGATCTTATTATTGTTTGTATCCAGAAGAGGGGGAAGAAAGCAATGAACCTCG 104021
QY 580 ----- 580
Db 104022 CTGGGAGAGGGGATTAACATATATCAGTTCTTCAGCAAAAGATCAGCATGTGTATAC 104081
QY 580 ----- 580
Db 104082 GTTTATCATCAACGACAGTTCGTTTATATATTAATTAATTAATTAATTAATTAATTCCT 104141
QY 580 ----- 580
Db 104142 CTCTAGACTTAGTTTAAAGGCTGTCTTGTGCGGCAATATAGCTCGGCTTATTTGCAC 104201
QY 580 ----- 580
Db 104202 GTACGAATCCGATGTGAAGAATTAATAAGCGTGCATTTTATTCATCCTTTTGTGA 104261
QY 580 ----- 580
Db 104262 AGTTATTCATTTTTTTTTTAAGTGTGTAACATTTTTTACTGGCAACATGTTCCATGAC 104321
QY 580 ----- 580
Db 104322 CTCGTACATCAACAATAATAGTACCTTACTTAAGTTAATTTTACCTTTTATGTAATAATTA 104381
QY 580 ----- 580
Db 104382 AGGAATACTTAATTAATAATTAATAATAAACAATCCGATTAAGATGATGTTATTTT 104441
QY 580 ----- 580

Db 104442 TGAACCTTCGTTAATAAACAACAACAACAAAGCTGTTATATAATATGTAGTTAAATATATAA 104501
QY 580 ----- 580
Db 104502 TTACTTTAACTGCTTGAAACCGTATATATACAAATTAAACGTTACATTTTTTAAGAACT 104561
QY 580 ----- 580
Db 104562 TAACAATAGCTAATAGCACCTTACTAATCAGTGAATAATAATATATATATATAA 104621
QY 580 ----- 580
Db 104622 TTATTAGTAAAAAGAAAATATTAACTAGTAAACATATATAGTCAGATGTTTAAAAAAAC 104681
QY 580 ----- 580
Db 104682 TGGCTAGTTTATCTCCACCTTTTGATTCATATTTTAAATTATTTGCCAGAAAACTAAC 104741
QY 580 ----- 580
Db 104742 ATGTTGACTTGTCCGGTTTAAGTGAAGAAATGGGTTATATGTTTCCAGCAATACACTT 104801
QY 580 ----- 580
Db 104802 TCAGACGGTGTCTTGTAGCGAAAATGCACAGACAGGTTACTTTTTCACAAACATGCGAG 104861
QY 580 ----- 580
Db 104862 AGAGGGCTTCGTCTTGCTGTTCCGCCGCCGATTTCTTGAACGAGGTTTCTCTACGCT 104921
QY 580 ----- 580
Db 104922 GCGAGGGCGCGCTTCTTAAGTGTATTACCAATATACCTATATGTTGCCATTTCTCTGCT 104981
QY 580 ----- 580
Db 104982 CAATGTCTCCATTATATCATTCTCTTTGTATCTGCGCATTTTGGCAAAATGCCCTGTC 105041
QY 580 ----- 580
Db 105042 CATGCTGCTGACGTGGGCTTTCGCGTTTAGAAGACAGAGTTGATTAAAGCTTACATG 105101
QY 580 ----- 580
Db 105102 CTTTTTTTCCCTCGCGTCATTCGCACTAGCTTGCCTGCTTCGGGACCGTGATCTGC 105161
QY 580 ----- 580
Db 105162 GTATAGATGATTTGCTCCTCAGTGTGAGCTCTCAGCAGTGAATATTAACACACACTT 105221
QY 580 ----- 580
Db 105222 TACTTAAACCCCATACAAAAAGCTTATCCAAATTGATTCTGCTTCACCTTCTCAGTCG 105281
QY 580 ----- 580
Db 105282 GCACATATCCAAAATGTTCCAGATGTAAGATTCGTTTCTAACAAAATTCATGTTGA 105341
QY 580 ----- 580
Db 105342 ACTTGAACCTGCGTAATCGCGCTAGTGTATATTCTAACAGATATGTAACTTGGCA 105401
QY 580 ----- 580
Db 105402 CATTACAGTATTTGTGTTTAAAGTGGTGACGTCACGTACTACGGCCGCTGGCAATGG 105461
QY 580 ----- 580
Db 105462 ACAACGAGTGGCACTGACACCGCGGTGACGGGTTGTATGTCATGTGACCGTCACAGCC 105521
QY 580 ----- 580
Db 105522 CTAGTTTAACTAT 105581

QY 580 ----- 580
Db 105582 ATATATATATATTCGAT 105641
QY 580 ----- 580
Db 105642 AATATGTCGACCCCTACTAGTGGAGTACTGTCTAACCTGCGCAAGCTGAATCTACC 105701
QY 580 ----- 580
Db 105702 CACATTTGCGGGTGGCGGTTTAAATGTAAGCCCTGTATACATTCACATGCT 105761
QY 580 ----- 580
Db 105762 CCTGTTAAATATGATTACAATTGACCTATGTTCTGATCAACAAATTAAAGCAAGCTTATTA 105821
QY 581 ----- Arg51u 582
Db 105822 TAGTACTAATATTTAAACAGACATTAAGTTATTTGAAGTCGCATGCTTTCCAGAGGGA 105881
QY 583 IlethrLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLys 602
Db 105882 ATCACACGAAAGCTGACATTTCCCTAAGCAGCCGCAAAATCCATTCCTAGAGATGGTGAAA 105941
QY 603 PheLeuLeuGluArgIleAlaProValHisIleAspTrpGluSerIle----- 618
Db 105942 TTCCTGCTGAGAGATTTGCACAGTCACATTCCTCTGAGGCAATTAAGTAATGAGA 106001
QY 618 ----- 618
Db 106002 ACTGTTTTTTTTTCTGTTGGAAAGAAAGTTTCATATAGCAGTCATCCGAGATGAAT 106061
QY 618 ----- 618
Db 106062 AAAGATTGTCTTAGCTTTTATTTATTTAGTTAATAGATTGGCAGCAACAAAAA 106121
QY 618 ----- 618
Db 106122 AAAAAAGATTGAAAATTTTATTTTCATTTAATGAAGATATTATCTGATGTTTAATTC 106181
QY 618 ----- 618
Db 106182 TTCAAGATTAATAAATAAAGCTTTCGTATACAGCAATCATTAATGTACTGTACATG 106241
QY 618 ----- 618
Db 106242 GACAAGATTAGATTAAGTACGACACACATTTGCTCATGAACCTCTTACAGTTATACAA 106301
QY 618 ----- 618
Db 106302 AACCAATGCAAGTACTTATGTGTGAACTAACAATTTGTTCACTTACAAAATTAAGTCT 106361
QY 618 ----- 618
Db 106362 TCAGTCAGTTTAAATCAAAATTTGTAATATATATACAGAGTGTAAATGACAGTTCTAT 106421
QY 618 ----- 618
Db 106422 GGAGTCTATCAGTCACCTCAGAAAGTGAAGTGAAGCAGTAATGAGAGAGAAAGAA 106481
QY 618 ----- 618
Db 106482 ACTGACTAGTGGGGAAGAAAGATTAAGCACTATATAGGGCCATGTCTTTTCCCTCCCTAAC 106541
QY 618 ----- 618
Db 106542 GGTGTGTAAACGCTCAATGCTTTTAAAGATGTGACATCTAGTGAGAAAAAGTGA 106601
QY 619 ----- SerAlaLeuI1 622
Db 106602 ACTGAGGTATTTGTGTGCTTTTGTGATTAACACATTATGTGTGTCTACAGTGCCTGTGT 106661

QY	622	elyslnlnValAsnlysserlleaspGlyThrAlaaspaspGluaspGluGlyValProth	642
Db	106662	TAAGCTGTTAAATTAAGTCAAATTGAAGGCCATCTGATGATGAAGATGAGGAGTAACTCC	106721
QY	642	rAspGlnAlaIleAglAglGlyLeuGluLeuLeuLys	654
Db	106722	TGATACCTGCCATCCGTCGCTGGACTTGAGCTACTTAAGGTAAATGTTTGCAATTGCAATGT	106781
QY	654	-----	654
Db	106782	AGTAGTTTTTATTTTAAATAAGACTTTTAAATACCATGTGGCGCATTTTTTCTTTCCCA	106841
QY	655	-----ValLeuserPheThrHisProIleSerPheHisSerAlaGluThrPheGlu	672
Db	106842	TGCGTCAGGTTTATCATTTTACCTACCCACCCATGCGTTCACCTCGGCTGAACATACGAGT	106901
QY	672	erLeuLeuAlaCysLeuLysMetAspAspGluCysValAlaGluAlaIleGluIleP	692
Db	106902	CTCTCTCTCAGTGTCTGAAATATGAGAGAGATAAATGGCTGAAGCAGCAATACAGATCT	106961
QY	692	helYasnThrGlySerLysIleGluGluAspPheProHisIleArgSer	708
Db	106962	TCAGAGACACACAGACAAAAAATTTGAGACAGAACTGCCCAAAATTCGCTGTTAGAACCC	107021
QY	708	-----	708
Db	107022	TCTCATATTCATTTACATMAAAACACCTCTACAGATCAGAACTGAAGCCTAAACTTA	107081
QY	708	-----	708
Db	107082	TTAAAAACCTGATACATGCTTTTAAATACAAATTAATGTAGTGAATGTTTTTTGCT	107141
QY	709	-----AlaLeuLeuProValLeuHisHisLysSerLysGlyPropio	723
Db	107142	AACTTTATGTGCAGAGACTATATCCCATACCTCATCAAGAAGCAAAAGAGAACCCCG	107201
QY	724	ArgGlnAlaLysTryAlaAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGln	743
Db	107202	CACCGAGCTTAACACAGCGGCTCCATTCATCCATGCCATCTTTCACAAATTAAGAGGTCCAG	107261
QY	744	PheAlaGlnIlePheGluPro	750
Db	107262	TTAGCTCAAAATTTTGGAGGTAAACATCTTTTTTGATGAAATATCACTGTATACAT	107321
QY	750	-----	750
Db	107322	TTCAATTCGCATATACAGTAACATATATAGAANAATGAAGGTCATGTTCCACATATTT	107381
QY	751	-----LeuHisLysSer	754
Db	107382	CACCACTGAATAATTTTCAGTCATTTTGTTGCTTCACAAATCAAGTAACCTTGAAAC	107441
QY	754	-----	754
Db	107442	ACTATTAATTACACACTGTGAATATTTTGGCTTGGCAATTAATTACTTGTGGCGTGTGG	107501
QY	754	-----	754
Db	107502	AGCAATTCACATAACCATATTTTCAAAATTAATAGTATATCTATTTCAGAGAAAAAATG	107561
QY	754	-----	754
Db	107562	TAATGAATGTCCTTTATTAAGAGGAAGGTCAATGCTGTGTTTCACATTTTATTAAT	107621
QY	754	-----	754
Db	107622	ATTGTTATATGATTAACCTTCTGAANAATGTTGTAANAATGAGTTGTTCCAAACACTGTT	107681
QY	754	-----	754
Db	107682	AAAGAACACTAATTTATATTAATATGTTTATTTAGAGGAGGAATGTAAAGAAATGCA	107741
QY	754	-----	754

D	b		107712	GCAATTCGATTGCAGGATGATCATCTGAATGANTCTAAATGCCTTTGAACTACGACTTA	107801
O	y		755	-----LeaspProSerAsn-Le	760
D	b		107802	AAACCCAAACTTGTGGAAAGAATCATATCACTACAGGTAAATAATGATCAGCGAATTAAT	107861
O	y		760	u-----	760
D	b		107862	CACCTCTAGAAGATCTTATTTTTTATTTTAGATCATCTTTTTCATCATGATGTTC	107921
O	y		760	-----	760
D	b		107922	TACAACAACAGCTTGTTCAGGCGATGAAGATTAAAGTAAGAGAGTGGCCAGCTCAATC	107981
O	y		760	-----	760
D	b		107982	GCCTTACTTCAACTTAATTGAACCTCATTTGGCTAAAATTAGATTGTGCCGATAGACGAT	108041
O	y		760	-----	760
D	b		108042	TGTATCAACGATAGGCAAAAATATCGCAGAGCGTAAACCTATGATGATTTAAAGAGC	108101
O	y		760	-----	760
D	b		108102	AATTTAGCTCCTTCGACGTGTAACATGTTTTAATGCATTATATAAATAATGAATG	108161
O	y		760	-----	760
D	b		108162	TATTACTATTTTAAATCATCATTAATAAAAAAATAAATAATATGATTCCTT	108221
O	y		760	-----	760
D	b		108222	CAGCTGTTTATATCAACATCACCTGCAGACACTTTTGGCAAATAATGATTTTCAAT	108281
O	y		760	-----	760
D	b		108282	TGCTCTTATCAAAAACTTTTACTGCCATCGAAAAATTTAGTCAAAATAGGCTAT	108341
O	y		760	-----	760
D	b		108342	TACTGACTGCCGCGACCATGTCTAAATGAATGAATATCATCATATGCGGATCTGAAA	108401
O	y		760	-----	760
D	b		108402	TTTTGTAATTCGCCAACACTAGCTGAATPACCTGTTCTAGGTGCCAGAAATTATTA	108461
O	y		760	-----	760
D	b		108462	AAAGCAATATTAATAATTCAGATAAATATATCAATATTTTTTCTCTAAACTGAA	108521
O	y		760	-----	760
D	b		108522	ACACTAATATATAAATTAATTAATGCACCCTTTCTATCTTGTTGATGACAGACAGT	108581
O	y		760	-----	760
D	b		108582	GTCCTGCTGTTCATAAATGAATGAATAATAGCAGCTTACTTACGCTATATTTTGC	108641
O	y		760	-----	760
D	b		108642	ACAGTAGATTGCTCCCTGCACCTGATGGTGAAGCAAGTGAAGTGTGACATCT	108701
O	y		760	-----	760
D	b		108702	ATTATGATCAATCTAATAATCTTCATAGAGTGGCAGACTGTTCTATATCAGCACACAA	108761
O	y		760	-----	760
D	b		108762	AAGCACATGTGGTGAAGTTTTTGAAGAAATGAATAATGGGTTGTACTACATTTGTTGS	108821
O	y		760	-----	760

Dh 108822 CTCGCAATTTCTACCAATGTATTTATGATGCAAGTTTACATTTCTGGTACATAT 108881
Qy 760 ----- 760
Db 108882 GAGACTAAAGTGTGTAAATTACAAACCTGACTATTTTGTATCTTGTGTAGATTAAG 108941
Qy 760 ----- 760
Db 108942 AACATGCAGACTTTATTTATGAACAATCAGTGCATCATTAATTAAGTGTCTTTGGCA 109001
Qy 760 ----- 760
Db 109002 CATCATATCTGAGGTGTGACGAGATTTTAAACTAGATATTGCAAAAAATAAATA 109061
Qy 760 ----- 760
Db 109062 AAAACAAAAGCAATTAACCCACTTTCCTGCAAAATATCAGCAGATTATATATAAATG 109121
Qy 760 ----- 760
Db 109122 GGTGCTAACTTGGCTTTTATGATGAATAATGTGACAAATGTATTTAAATGTTTAA 109181
Qy 760 ----- 760
Db 109182 TTTATTTTATTTTATTTATGCAATTTGTTTCTGACGAGTTTTCAAAACATTTTA 109241
Qy 761 ----- GluHisIle 763
Db 109242 TCTCTGTGTTTGTCCACGCCCTGTCTCGAGCTGTAATGCAGATGTACCTGAAACAGCT 109301
Qy 763 uIleThrProLeuValIhrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaI 783
Db 109302 GATCACTGCACGTCGTCTCTGCGAGACATCTCCATGCTGGGCCCGGAGCATTTGGCATC 109361
Qy 783 aProTrrpysSerTrpValAlaIhrPheIleValIlysAspLeuMetAsnAspArg-- 802
Db 109362 TCCCATGAAGTCAAATTGTAGCTAAATTTATGTAAAGACCTTGTGATGAATGACAGAGT 109421
Qy 802 ----- 802
Db 109422 GAGCATATGTGATGTATTATTCCTTTCAGCATGTGGAATGTTTTTATAGCCCTTGCAT 109481
Qy 803 -----LeuProGlyIysIsthTrhIleLeuTrpVal 813
Db 109482 GTTAATTTTACCTCCTCATCTCTTAAGTCTGTG- GGAACAACAAAATGGCGGCTCTGAGC 109540
Qy 813 lProAspGluGluValSerProGluTrhMetValIys----- 825
Db 109541 AGCTGATGATGAGGTTCCTCGAAGTTTGGCTAAGGTAAATATATATCATATA 109600
Qy 825 ----- 825
Db 109601 AAATTTCACTTGTCCCTTTTAAAAATTCGCCAAGACACTTTCGACTGAGAAAATACAGTT 109660
Qy 825 ----- 825
Db 109661 AATACATAATTAAGGTAGAAAACGTCAATTTTGTATGTGATTTATTTATTTGTTGTCT 109720
Qy 826 -----IleGlnAlaIleIysMetMetValArgTrpLeuGluIymetIysAsnAsn 843
Db 109721 GTTTCAGGTGCACACCACTTAAGCTGTGTGTCGCCGTGTGTAGGAATGAAGAATAATC 109780
Qy 843 lsserIysSerGlyThrSerThrIleuArgLeuLeuThrTrhIleuHisSerAspGlyA 863
Db 109781 AATCCAAATACAGCTACTCTCTCAGACAGCTGCTCTCAGCCATGCTGTCACTGAGAGAG 109840
Qy 863 sPLeuThrGluGlnIlyIysIle----- 870
Db 109841 AACTCAGGAGAGAGAAGAAGTTAGTACTCAGTACTAGTATAGTATGAT 109900
Qy 871 -----SerIysProAsp 874
Db 109901 AATAAAGACATATGCTCTTAACACACATTTGTATGTTTGTGTACACAGCAAGTCTGAC 109960

Qy 875 MetSerArgLeuArgLeuAlaIaGlySerAlaIleValIysLeuAlaGlnIuProCys 894
Db 109961 ATGTCTCGACGTAAAGCTTTGACACCGGAGATGCCATCTGCAAACTGACACAAAGCCCTGT 110020
Qy 895 TyrHisGluIleIleThrLeuGluGlnIyrGlnLeuCysAlaLeuAlaIleAsn----- 912
Db 110021 TACCATGACATATTACTCCAGAACAGATTCCAGCTCTGTGGACTGTCTATTAATGTGGGT 110080
Qy 912 ----- 912
Db 110081 TCAGCTTACTGGGCAATATATCACTCAATATTTAACTGAAGAACAGATAGCCTTGAC 110140
Qy 912 ----- 912
Db 110141 TTTTAGGACGACAAAAACAAATTTCCCTTTCACACATGATGACGCTTAATTCAGGGTC 110200
Qy 912 ----- 912
Db 110201 TGTGCACACATCAATATAGCAGGACATTTTGTTCAGATGTTTGTATGATTAATGGCTGAA 110260
Qy 912 ----- 912
Db 110261 GATCAATTTTAAAGTAGACTTTTCTTTTCTTTTGACAGATCTTTAGATTTAGAC 110320
Qy 912 ----- 912
Db 110321 TAGACCTGTGTACACACACACACAAAGCAATTAACAAAAAAACTCTTAACTAAA 110380
Qy 912 ----- 912
Db 110381 TTGTAATAAAGAAATGCTGTGTATATGCGCTATTTTGACACCTGTAGAGTAATGGCTAAA 110440
Qy 912 ----- 912
Db 110441 TGTGTCTCTTGTGACTACATGTCTACAGCAACACTCTTCATTTATAGCTTTCTCT 110500
Qy 913 -----AspGluCysTrpGlnValArgGlnValPheAlaIleIlyIle 926
Db 110501 CTTTCCCTCTCTTATAGATGAGTGTCTATAGGTGCGACAGATCTATGCTCAAGAGTT 110560
Qy 926 uHisIysGlyLeuSerArgLeuArgLeuProLeuGluTrpMetAlaIleCysAlaLeuCy 946
Db 110561 ACATGTGCTGTGTGAAGTGTGTGTACCATGAGTACAGGCTGTATTTGGCTTTAG 110620
Qy 946 sAlaIlysAspProValIysGluArgArgAlaHisAlaArgGlnCysLeuValIysAsnI 966
Db 110621 TGCTAAAGACCCAGTAAAGAGCAGGCTCATGTCTGACAGTGCCTGTGAAGAACAT 110680
Qy 966 eAsnValArgArgGluTrpLeuIysGlnHis-Ala----- 977
Db 110681 CAGTGTCCGACAGAGTACATCAACAGAACCTTATGGCTCATGTGATACTTTTGT 110740
Qy 977 ----- 977
Db 110741 TTAATTTTAAACATATATGTTTGTACTGTATACACATACACACACCGGCAAT 110800
Qy 978 -----AlaValSerG 981
Db 110801 TTCTCGGAGAGCTGTGTAATGTAAAGATCTATCTAATCTTTTGTTCATTTTCAG 110860
Qy 981 lulysLeuLeuSerLeuLeuProGluTrpValValProTyrTrhIleHisLeuAlaIh 1001
Db 110861 AAAAGCTCTGTCTCTCTCCAGAAATATGTGTGCCATACATGTGATCTTCTAGCTC 110920
Qy 1001 lAspProAspTrpValIysValGlnAspIleGlnIleuIysAspValIysGlu----- 1019
Db 110921 ATGATCCAGACTTACTTAACACACAGATTTGGAGACACTTGAAGATGTCAAGA- GTAC 110979
Qy 1019 ----- 1019
Db 110980 GAGATCATATTTTGGCATTTTCTTTCACCTCAATGAGCGCTAATTAAGTCAATCT 111039

Db 22924 -ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCCGATCAATGACAGCTTGT 22981
 QY 270 euleSerValIleuProGlnLeuGluPheLysLeuSerAsnAspAsnGluGluArgL 290
 Db 22982 TGTGTACGTCTTACTTACCTGAGTGAACAACACTGTCTGACGAGAGATCTGAGAGGC 23041
 QY 290 eu----- 290
 Db 23042 TAAAGTATATTTATGAGCATGTACTTATTTGCAATATTAACATATTTTCGTTA 23101
 QY 291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAla 309
 Db 23102 CTACAGAGGCAACCACTCTCTTGTCTGATGTTCTCGAAAGAGCTCGCAACTATCGCA 23161
 QY 309 ergLAsnLysProLeuTrpGlnCysTrpLeuGlyArgPheAsnAspIleHisValProI 329
 Db 23162 AAAAGTACCCCAACCTGTGTGAANAATCTTCTTGGGGCTTGTGGACATCATCCGACCG 23221
 QY 329 leaTrgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAla 349
 Db 23222 TTCCATCAACAGTGTGTTCAGTCATCATGCACTTCCTCAATCACCACCAAGTCTTGACG 23281
 QY 349 ysaAspLeuThrGluTrpLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
 Db 23282 ACGATATCACCGAATAATTCGGCTTCGAATTCAGATCTTGACGAAGTGSTCCGCCACG 23341
 QY 369 spValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp---- 387
 Db 23342 AGGTGTAAAGCTATTGTGGAACCTCGCAAGCGCGACTTCACCTCGTTCTCGAAGCGC 23401
 QY 388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
 Db 23402 CCGATCTACAGAAATGTGTGGCGGAGCGGAGCTAGATAGAACATACAAATTTGCGAGCG 23461
 QY 407 IuaIleMetMetGlyLeuAlaGlnIleTrpLysLysTyraLeuGlnSerAlaIleArgL 427
 Db 23462 ATGCGATGAATGTGTCTGGCTCTACATCTACAAGCGCGCAATTTGGCAACCCACAGATCTAA 23521
 QY 427 ysaAspAlaAlaLys--GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTrpTyrg 446
 Db 23522 GCACTGGCTTAAAGTCAGGGGTGACTGATTAAGAACAAAGATCATGATGATTAACA 23581
 QY 446 lnaAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTrpMetValProH 466
 Db 23582 AAGTGGCTGTGGAGAGCGGCTGTGTGGAACGCTTATCATACCTGCTGCTGCTCCT 23641
 QY 466 isaLeuGluTrpThrGluArgMetLysCysLeuTrpTyraLeuTrpAlaThrLeuAspL 486
 Db 23642 ATTAACTAGTCTCCGAGAGCGCATGAAGAAGCTATATCATTTCTAGCGCATCTCGATG 23701
 QY 486 euaAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
 Db 23702 CCAATGCGACCAAGCTTGTGTGTCGAGTCGAGAAGAACAAATGAAGACGCAATACGG 23761
 QY 506 allyAspLeuLeuAspLeuLysGln-----ProLysThrAspAlaSerV 522
 Db 23762 TCAGTGTATGTGATCAACTACACCACTTAAGAGATTCAACGCCGGGTACTGTAGTACG 23821
 QY 522 allyAla----- 524
 Db 23822 TCAGCGCCAAAGCAGGCAACATTGCAAGTGGGTTCTAATAAAGTTTCTTTTGATAGT 23881
 QY 525 --IlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGln 543
 Db 23882 TTGTAAACAGAGTGTGTTTTTTTCTTCTAGACTGCTCCAGATCCCTAAAGCGGCA 23941
 QY 544 AspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeu 563
 Db 23942 GAGTACCTAACCCAGTTTACGACACACACCTGCGAAGAAAGATGCTGACGCTACGATGAT 24001
 QY 564 GluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgLLeu 583
 Db 24002 AACATGTCTTAAACGCGACGATAGCTGTGCGAGGTGCTGACACGATGGAGTCTC 24061

QY 584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetLysPhe 603
 Db 24062 CTGAAAAGCTGTGGCGCCACAGCTCCAA---TCGAATCTGATTTACACACAGTTAAAGATG 24118
 QY 604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuLys 623
 Db 24119 CTGATGAGCGCGGTGATGATGATGTGACAAAGGATCTATTTGGCTGATTTAGG 24178
 QY 624 -----Gln 624
 Db 24179 TAAATGCTTTTAAAGTCCCTTTTGCACACTTAATCCGTAATCTTTTGTCCAGCTTA 24238
 QY 625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluValProThrAspGln 644
 Db 24239 ATTGACAAATGCAATCGAAGAAAGGTCCATGTGTGAAGATGCGAATTTGGCTCAGGAA 24298
 QY 645 AlaIleArgAlaGlyLeuGluLeu----- 653
 Db 24299 GCGGGCAGCGCGGACATCAAGTTGTGGCCGTGATAGAGATAATTAAAGTTATGATA 24358
 QY 654 -----LysValLeuSerPheThrHisProIleSer 663
 Db 24359 CAGAAATCCATTTGATGCTTTATACTTGGCAGATG-CTTTCATACGCTTTTGGCGTCAT 24417
 QY 664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys 683
 Db 24418 TTCTTACCGGATCTGTGCTGCTGCATTTAATTTCCCTGCTCAGCTTACGACAGAGATTA 24477
 QY 684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700
 Db 24478 GTTCCACACCTGCTGTGAACACACTAATCATATGATGGCGGCTATGACGCCACTGATGAC 24537
 QY 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
 Db 24538 GAACCTACACCGGCTATTTTCGAGAGATGCTGCTGGTGTGCGAAGACTTCGCACTTAT 24597
 QY 721 GlyProProArgGlnAlaLysTyraAlaIleHisCysIleHis----- 734
 Db 24598 GGAACCCCAAGCAAGCAAGACGCGGTGATGTATTTTCTGTAACACGCAAGTCTGCG 24657
 QY 735 -----AlaIlePheSerLysGluThrGln----- 743
 Db 24658 GCTTTCACGTATGGAGCAACAAGTGAAGTGAAGTGCCTCCACACTACACAAACAGTGC 24717
 QY 744 -----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
 Db 24718 CATGCTATTTTACAGAGATCATGAGACGCTTGCCTTAACATGACACCA---AAGTGT 24774
 QY 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
 Db 24775 GAGCATACGCGCAACAAGATTGTGACTGTGGTCACTTGCCTCAATATATGCCACAGGCC 24834
 QY 781 PheAlaIleProTrpLysSerTrpValAlaThrPheIleValLysAspLeuMetAsn 800
 Db 24835 TTCTTAAGCCCAATTAAGATGATGATGCGGACGAATTTGCAAGAAGCTGCTTATCAAA 24894
 QY 801 AspArgLeuProGlyLysLysThrThrLysLeu-----TrpValProAsp 815
 Db 24895 GAA---GTTCCTGGCGACGCGAGCATAGCAACTGCCAGACAGCAAGTATGGTGTGCCAA 24951
 QY 816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
 Db 24952 GAGAACTACCGCGGACACACTGTATGAAGTGTGATGCGCTCAAGGCTTATAGCGCAGGTGG 25011
 QY 836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrIleAlaGlyLeuThr 855
 Db 25012 CTTTGGGATTTGCTTACCGAT---GACACAGCTCCCGCAAGAAACATTTCCGAATCGAGTGG 25068
 QY 856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
 Db 25069 GCGTTTGTTAACACAGAGGCGATTTGCTGCGCCAGAACCGTCTTTCGCGTCCGCAAGAA 25128

OY	876	serirgyleuargyleualalagllyserAla1lleVallyseuilaclngluyProcytyr	895
Db	25129	TCTTGGCTGGCGCCCTGGGGGACCCCTGCCCATCTCCAGTGGCGCAAAAGGGCGTA	25188
OY	896	HisglullellethrlheugluginTy--GlnleucysAlauehlaileasnsprGlu	914
Db	25189	GGTGATCAGTAACAGCGCGTAGCAGATATTTCACACTTCCACGCTGATGGTAGTAGAC	25248
OY	914	-----	914
Db	25249	ATTCAATCTTGCTTAGACCTTAGTCAATGTCTATCTCGGTGTTCCGGTAGCGTATC	25308
OY	915	---CysTyrglnAlaArgGlnValRheAlaGlnlyLeuHislysglyLeuSerArgLeu	933
Db	25309	CGGTGCCA-GAAGTTCGGGAAATCTTTGCTCGGCAACCTGGCACAAAGATTAAAGCAGAGCT	25367
OY	934	Arg-----LeuProleuGluTyMetAla1leGysAlaLeuGysAla-----	947
Db	25368	TTGGCCAGGAACGTGTTGGCGGTGGACTGTATGGCTTGATGTGCTGGCTGGTCTAGAG	25427
OY	948	-----LysAspProVallysglyArgArgAlaHisAlaArg-GlnCysLeuVally	964
Db	25428	ACTGAGAGAAAGTAGTAGTTCAGAAACATCCGATCCGGTGGAGAAAGTATTGTATAA	25487
OY	964	S-----AsnIleAsnValAr	969
Db	25488	TATTTTCAACAGATTCGACAGACCTTGCTGCTCACTATGAGAAACGAGATTAACAAACG	25547
OY	969	gArgGluTyLeuLys-----	974
Db	25548	CGGGGAATATCTCAAGACCTGGCTATGACATGAGTGTATTACTTTTGGAAATCATTT	25607
OY	975	-----GlnHis-AlaAlaValSerGluTyL	983
Db	25608	GTGATTATTAAACCTTAATCAATTAATTATTCACGCTCCGACAGCTCAACGAGTCAACAT	25667
OY	983	euleuSerleuLeuProGluTyValValProTyThrIleHisLeuLeuAlaHisAsp	1003
Db	25668	CATTACACATATACCTCGACTACATAGGTGGCTTTCCTCTCCCGGTGGTGCACAGATC	25727
OY	1003	roAspTyrVallyValGlnAsp1leGluGlnLeuLysAspVallysglyCysLeuTrp	1023
Db	25728	CACGCTTCACAGAAATCACAGAGACTACGTACAGCTGGCGCAAGATGGGAAAGTGCCTCGTT	25787
OY	1023	heValleuGluLeuMetAlaLysAsnGlnAsnSerHisAlaPheIleArgTyrLysM	1043
Db	25788	TCATTTCTGGAGCGCGCTATGGCCAAACGAAACGTTTGGCATACCTCTACAAAGCAC	25847
OY	1043	etValGlnAsnIleLysGlnThrLysAspAlaGlnIleProAspAspAlaLysMetAsn-	1062
Db	25848	TGCTGACAGCTGAATAAAGCATCGGAGTTCAGTCTGGGG--TCGGCAAGCGCGCAACT	25904
OY	1062	-----	1062
Db	25905	ATGATATCTCACTCAATCATTTTCCAGGCTTTCCCTTAAATTAATTATTATACATATAT	25964
OY	1063	-----GluLysLeuTyTrhValLysAspAla1laMetAsnIleIleMetSerL	1079
Db	25965	TTTTTCGGTATACGAAAAATGTGGCGGCTGTGGCATCTTCATATGATTAATTCAGCTCA	26024
OY	1079	ys-----SerThrThyTyrSerLeuGluSerProLysAspProv	1092
Db	26025	AGTTCACTCATTTTGATGGCAACACAGACACTTTTCCATTCGCGCGG-----G	26072
OY	1092	alleuProAlaArgPhePheThrGlnProAsp--LysAsnPheSerAsnThrLysAsnT	1111
Db	26073	CTTTGCCAGAAATGTATTATTAAGACCTGGCGCTGGCAATTTCCAAAACAATGACGCTCT	26132
OY	1111	tyrLeuProProGluMetLysSerPhePheThrProGluTyProLysProLysThrThrAsnValL	1131
Db	26133	ATATACCGCTGACGCGT-----TATACGGGTGGAGCCAAATTCACAGCAAAAGCG	26183
OY	1131	eueGlyAlaValAsnLysProLeu---SerSerAlaGluTySerGlnSerGlnThrTySers	1150

Db	26184	CCGCACAGC-AATGACACACTCGGACGACACATGTCGCT-CCAAAGACCGCGGGAACA	2624
Qy	1150	erArg-----MetGluThrValSerAsnAlaSerSerS	1161
Db	26242	GTCAATCATGATGATGATAAAATCCGACAGTCAGTAGAGAGTCTCTGTGGAAAATATCTGAGATG	2630
Qy	1161	erSerAspProSerSerProGlyArgIleYsgIyArgLeuAspSerSerGluMetAsp	1181
Db	26302	AATCTAACTAGTGTGTGAAATTTGATTTACAGACAGCAATCTGTTCGACCAACATATC----	26355
Qy	1181	IsSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysSerSerAspLys	1200
Db	26358	-----GAGCGCTGACACTACGACGCCATGGCCAAACAGACGCGCAGAGG	26405
RESULT 15			
AC007474			
LOCUS	AC007474		
DEFINITION	Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone		
ACCESSION	BACR14119, complete sequence.		
VERSION	AC007474		
KEYWORDS	AC007474.5 GI:13162477		
SOURCE	HTG.		
ORGANISM	Drosophila melanogaster (fruit fly)		
REFERENCE	Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 192763)		
	Celniker,S.E.; Adams,M.D.; Kronmiller,B.; Tyler,D.; Wan,K.H.; Holt,R.A.; Evans,C.A.; Gocayne,J.D.; Amaratunga,P.G.; Brandon,R.C.; Rogers,Y.; An,H.; Baldwin,D.; Banazon,J.; Beeson,K.Y.; Busam,D.A.; Carlson,J.W.; Center,A.; Champs,M.; Davenport,L.B.; Dietz,S.M.; Dodson,K.; Dorsett,V.; Doup,L.E.; Doyle,C.; Dresnek,D.; Farfan,D.; Ferreira,S.; Houck,J.; Hoskins,R.A.; Hostin,D.; Howland,T.J.; Ibegwan,C.; Jallali,M.; Kruse,D.; Li,P.; Mattei,B.; Moshrefi,A.; McIntosh,T.C.; Moy,M.; Murphy,B.; Nelson,C.; Nelson,K.A.; Nunoo,J.; Pachtel,J.; Paragas,V.; Park,S.; Patel,S.; Pfeiffer,B.; Phoumenavong,S.; Pittman,G.S.; Puri,V.; Richards,S.; Scheeler,F.; Stepleton,M.; Strong,R.; Svrtkas,R.; Tector,C.; Williams,S.M.; Zaveri,J.S.; Smith,H.O.; Rubin,G.M. and Venter,J.C.		
JOURNAL	Sequencing of Drosophila chromosome 2R, region 49A-49B		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 192763)		
	Celniker,S.E.; Agbayan,A.; Arcaina,T.F.; Baxter,E.; Blazej,R.G.; Buehloff,C.; Champ,M.; Chavez,C.; Chew,M.; Ciesiolka,L.; Doyle,C.M.; Farfan,D.E.; Galle,R.; George,R.A.; Harris,N.L.; Hoskins,R.A.; Houston,K.A.; Hummasti,S.R.; Karra,K.; Kearney,L.; Kim,E.; Lee,B.; Lewis,S.; Li,P.; Lomoton,M.A.; Mazda,P.; Moshrefi,A.R.; Moshrefi,M.; Nixon,K.; Pachtel,J.M.; Park,S.; Pfeiffer,B.; Poon,L.; Sequeira,A.; Sethi,H.; Snir,E.; Svrtkas,R.R.; Wan,K.H.; Weinburg,T.; Zhang,R.; Zieran,L.L. and Rubin,G.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-MAY-1999) Drosophila genome center, Lawrence Berkeley		
TITLE	Laboratory, MS 64-121, Berkeley, CA 94720, USA		
COMMENT	On Feb 28, 2001 this sequence version replaced gi:5670617.		
	Sequence submitted by:		
	Berkeley Drosophila Genome Project		
	Lawrence Berkeley National Laboratory, MS 64-121		
	Berkeley, CA 94720		
	This sequence was assembled using end sequences from a whole genome		
	shotgun and from subclones of this BAC and its neighboring clones.		
	For further information about this sequence, including its location		
	and relationship to other sequences, please visit our sequence		
	archive web site (http://www.fruitfly.org/sequence/) or send email		
	to bdgpf@fruitfly.berkeley.edu .		
	Location/Qualifiers		
	1. 192763		
FEATURES	/organism="Drosophila melanogaster"		
SOURCE	/mol_type="genomic DNA"		
	/strain="Y; cn bw sp"		

```

/db_xref="taxon:7227"
/chromosome="2R"
/map="49a-49b"
/clone="BACR14L19 (D591)"
/clone.lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcORI in
PBACE3.6)"
BASE COUNT      54505 a 41459 c 42141 g 54658 t
ORIGIN
Alignment Scores:
Pred. No.:      2,56e-80      length:      192763
Score:          1729.50      Matches:      463
Percent Similarity: 49.43%      Conservative: 270
Best Local Similarity: 31.22%      Mismatches: 436
Query Match:     24.04%      Indels:      318
DB:              3          Gaps:      37

US-09-512-581b-2 (1-1391) x AC007474 (1-192763)
QY      13  ILeThrTYrProProGlyValLysGluIleSerAspLysIleSerLysGluMetVal 32
      146305 ATAGGTACCCACACCGGATGCGCGCGCTGTGAGATTTGGCAGACAGACGACGATA 146364
QY      33  ArgArgLeuLysMet-ValVal-----
      146365 AGGCGCCTCAAGGTGAGTCTGCTCCCTGCTTACCAAACTTCTCACTTAGAATGA 146424
QY      40  -----LysThrPheMetAspMetAspGlnAspSerGluGlu 52
      146425 TTCTGTTTACAGACCTCCCAATGTTCTGCAGACCATGAGACGAGACGACAATCTCTA 146484
QY      52  uLysGluLeuTYrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
      146485 CCAGCAA---TACATACCATTTGGCGCCCTCCATCTGCTGAGACACTTTTATGCAACATCC 146541
QY      72  oGlyLysAspValArgLeuLeuValAlaCysLysLeuAlaAspIlePheArgIleTYrAl 92
      146542 ATCCCGTGATGTCAGCTCTTAACGCGCTGCTGCGCATGCTCGCGGTTTATGC 146601
QY      92  aProGluAlaProTYrThrSerProAspLysLeuLysAspIlePheMetPheIleThrAr 112
      146602 GCCAGAGGCTCCCTACAAAGACGACGATCAATGCAATATATTTAAGTTTATTATCAA 146661
QY      112  gGlnLeuLysGlyLeuGlnAspThrLysSerProGlnPheAsnArgTYrPheTYrLeuLe 132
      146662 GCAACTGCAGGCTCGAAGGATCGCGCATCCCTCTTCAAGCGTTACTTCTATGCT 146721
QY      132  uGluAsnIleAlaTYrValLysSerTYrAsnIleCysPheGluLeuGlnAspSerAsnGl 152
      146722 GGAAGAACCTTCTTGTCTCAAGCTTTTAATATGTTTTCGAACCTGGAGGACTGCCAAGA 146781
QY      152  uIlePheThrGlnLeuTYrArgThrIlePheSerValIle-----
      146782 AATCTTCCAAGACCTGTTTGTAGTACCATCTTCAGATTGTCAGTAATCGAGCTTCTCT 146841
QY      166  -----AsnAsnGlnHisAsnGl 171
      146842 TTATCTCGCACTTTATATACATGTTATATATTTTCTAGTACACGACGACATGT 146901
QY      171  nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluLysAspThrVa 191
      146902 CAAGGTCTACTACTTTTCTTGTAGTGTGTAAGCCCTTGATATCAACAGCCGCAATCT 146961
QY      191  lSerGlnGluLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
      146962 GTGGGCGGAATGTTGATGATCTCATCTGATTAAATGTAAGCCGTAACAAATCTAACAA 147021
QY      211  nLysGlnAlaTYrAspLeuAlaLysAlaLeuLeuLysArgTYrHisLagIleAlaIleGlu-P 231
      147022 CAAGTTTGCTGCACAATTAACAGAACAACTTCTCACAATAAAGCGGAGTCTCGAGTCC 147081
QY      231  ro-----TYrIleThr-----

```

```

      147082 CACCATCAAAATGTAATATTTGCTGATCTTTGGTCACACAGCTTATTCACCTAATCC 147141
QY      236  -----PhePheAsnGlnValLeuMetLeuGlyLys-----ThrSerIleSera 250
      147142 CTGATCTACAGCTTTTATACGCGCTGTGTGTGATGATTAAGCCAAACAAACTGTCC- 147200
QY      250  spleuSerGlnHisValPheAspLeuIleLeuGlnLeuTYrAsnIleAspSerHisLeuL 270
      147201 -ATTAAACAAACAAATATGATATCATTTACAGGCTTAACCCGATCAATGACGGCTTGT 147258
QY      270  euleuSerValLeuProGlnLeuGlnLupheLysLeuLysSerAsnAspAsnGluArgL 290
      147259 TGTGTGAGTCACTCACTGAGAACAACTGTCTCCAGAGATGTCGAGAGGC 147318
QY      290  eu-----
      147319 TAACTGAGTAATTTATACAGCATGACCTCATTTTGCATATTAATTAACATATTTTCGTTA 147378
QY      291  -----GlnValValLysLeuLeuAlaLysMetPheGluAlaLysAspSerGluLeuAla 309
      147379 CTACAGAGGCAACACACTCTGCTGCTGATGTTCTCGAAAGAGACTCGCAACTACGGA 147438
QY      309  erGlnAsnLysProLeuTYrGlnCysTYrLeuGlnArgPheAsnAspIleHisValProI 329
      147439 AAAAGTACCCCAACCTGTAATAATCTTTGGCGCTTCTGCGACATCACCGAACAG 147498
QY      329  leArgLeuGlnCysValLysPheAlaSerHisCysLysMetAsnHisProAspLeuAla 349
      147499 TTGCGATCAAACTGTGTTCACTCATCTGCTGCTCAATCACTCACTCACTCACTCACT 147558
QY      349  ysAspLeuThrGluTYrLeuLysValArgSerHisAspProGluAlaIleArgHisA 369
      147559 ACGTATCACAGCAAAATGTGGCGCTCGAATATCAGATCTTACAGAGTGTCCGCCACG 147618
QY      369  spValIleValSerIleValIleThAlaAlaLysLysAspIleLeuLeuValAsnAsp- 387
      147619 AGGGTGAATAGGCTATGTTGTGAACCTGCCAAGCGCGCACTTCACCTGTTCTGCAACCC 147678
QY      388  -HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTYrArgValArgLysG 407
      147679 CCGATCTACTAGAAATGTGTGCGCGCAGCGGACGCTAGATTAAGAAATGCAAAATTCGCA 147738
QY      407  lAlaMetMetGlyLeuAlaGlnIleTYrLysLysTYrAlaLeuGlnSerAlaAlaGlyL 427
      147739 ATCGATGATAGTGTGCTGCTTACTATCAAGCGCGCAATTTGCAACCAAGATCTAA 147798
QY      427  ysAspAlaAlaLys-----GlnIleAlaTYrIleLysAspLysLeuLeuHisIleTYrG 446
      147799 GCACGTGCTTAAGCTGAGGCTGACGTGAATTAAGAACAGATACATGATGATACAA 147858
QY      446  lAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTYrMetValProH 466
      147859 AAGTGGCTTGGAGACCGCTGCTTGTGAAGCGCTACTTATTCACCTGCTGCTGCTCT 147918
QY      466  lAsnLeuGlnTYrThrGluArgMetLysCysLeuTYrTYrLeuTYrAlaThrLeuAspL 486
      147919 ATAACTAGCTCCGAGAGACGCGCATGAAGCATATATCAATTTGCTAGGAGATCTGATG 147978
QY      486  euAsnAlaValLysAlaLeuAsnGlnMetTYrLysCysGlnAsnLeuLeuArgHisGlnV 506
      147979 CCAATGCGCAACAGGCTTGTGTGAGCTGAGAGAACCAAAATGAACACGCGAAATACGG 148038
QY      506  aLysAspLeuLeuAspLeuIleLysGln-----ProLysThrAspAlaSerV 522
      148039 TCACTGATTTGGATCAAGCTACACCACTTAAGAGATTCAACGCGCGCTTACTTAGTACG 148098
QY      522  aLysAla-----
      148099 TCAGCGCAACAGCGCAACATTTGCCAAGTGGGTTATATAAGTTTCTTTGATAGT 148158
QY      525  --lIlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
      525  ---lIlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543

```

D 148159 TTGTAACCACTGATGTTTTTTTTTTCTTGACACTGCTCCAGATCCCTTAACGCGCA 148218
Q 544 AspPheMetLysLysPheThrGlnValLeuGluAspArgLysIleArgLysGlnLeu 563
D 148219 GAGACCTTAACCCAGTTTACACAACTCGGAAAAATGCTCGCTCAAGATGCAATT 148278
Q 564 GluValLeuValSerProThrCysSerCysGlnIleGluGlyCysValArgLysIle 583
D 148279 AACATTGCTCTTAACCCGACGATAGCTGCGAGAGTGTGTCACACAGATGGAGTTCTC 148338
Q 584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603
D 148339 CTGAAAAAGCTTGCGCCCGCCAGCTCCAA---TCGAACTGTATATACAAACAGTAAAGATG 148395
Q 604 LeuLeuGluArgLysIleArgProValHisIleAspThrLysSerIleSerAlaLeuIleLys 623
D 148396 CTGATTGAGCGCGGCGATGATGATGAGCAAGAGTCTATTGGCGTACTATTAGG 148455
Q 624 -----Gln 624
D 148456 TAAATTGCTTTTAAGTCCCTTTTGCACACTTAATCCGTAATCTTTTGCACGCTA 148515
Q 625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGlyValProThrAspGln 644
D 148516 ATTGACAAATGATCGAAGAAAGGCTCATGTGTAAGAGATCGGAATTTGCGCTCAGGAA 148575
Q 645 AlaIleArgLysLeuGluLeu----- 653
D 148576 GCGGGCGAGCGGCGACTCAAGTTGTTGCGCTGAGTATAGAGATTAATTAAATTATGATA 148635
Q 654 -----LysValLeuSerPheThrHisProIleSer 663
D 148636 CAGAAATCCATTGATGCTTTTATATTGACAGATG---CTTTCATACGCTTTTGGCGCAT 148694
Q 664 PheHisSerIleArgLysPheGluSerLeuLeuAlaLysLeuLysMetAspAspGluLys 683
D 148695 TTCTCTACCGATCTGCTGCGCTCATTTAATTTCCCTGCTCAAGTACGACAGATATAT 148754
Q 684 ValAlaGluAlaIleLeuGlnIlePheLysAsnThrGly-----SerLysIleGlu 700
D 148755 GTTGCACACACTGCGCTCAAGACACTAATCATGTGGGGCGGTATGACCCACTATTTGAC 148814
Q 701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisLysSerLysLys 720
D 148815 GACCTTACACCGGTATTTCTGACAGAGCTGCTCGGTGCGCAAGACTTGCGACTATTT 148874
Q 721 GlyProArgGlnAlaLysTyrAlaIleHisCysIleHis----- 734
D 148875 GGAACCCCAAGCAAGCAAGCAAGCGGTGATGTTTTCGTAAACAGCCACTGCTGC 148934
Q 735 -----AlaIlePheSerSerLysGluThrGln----- 743
D 148935 GCTTCCACTGATGGACCAACAGTGGAGTGGAGTGCCTCCACACTACACAAACAGTGC 148994
Q 744 -----PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
D 148995 CATCTCTATTTTCAACGAGATCATGACAGACCTTGCCTTAAACAGTACACCA---AACTGT 149051
Q 761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuAlaProAspGln 780
D 149052 GACATCATGCGCCAAAGATTTGACCTTGAGTACATTCGCTTACAAATATGCCACAGGCC 149111
Q 781 PheAlaAlaProThrLysSerTyrValAlaIleThrPheIleValLysAspLeuMetAsn 800
D 149112 TTCTCTAAGCCCAATTAAAGATATGATGCGCGACGATTTGTCAAAGAGCTGCTTATCCA 149171
Q 801 AspArgLeuProGlyLysLysThrThrLysLeu-----TyrValProAsp 815
D 149172 GAA---GTTCTGCGGACGACAGACTACGAACTGCGCAGAGACAGATGATGTTGTCGCCAA 149228
Q 816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTyr 835
D 149229 GAGAAACTACCGCGGACACTTATGTAACCTGGATCGCTCAAGGCTATGTCGCCAGGTGG 149288

Q 836 LeuLeuGluMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuThr 855
D 149289 CTTTGGGATTTGCGTACCAGT---GACACCGCTCCGCAAGAAATTTCCGATGCTGGCG 149345
Q 856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
D 149346 GCGTTTGTATACCAACGAGGCGATTTGCTCGCGCAAGAAACGCTTTGGCGTCCGAGAAA 149405
Q 876 SerArgLeuArgLeuAlaArgLysSerAlaIleValLysLeuAlaGlnLysProCysTyr 895
D 149406 TCTTGGCTGGCGCCCGGCGCGAGCTGCGCAAGAGTGTGTCGAAAGTGTGTCGAAAGGCGCTA 149465
Q 896 HisGluIleIleThrLeuGluGlnTyr---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
D 149466 GGTATCATGATACAGCGCTGACAGATATTGCGAGCTTTCCACAGCTGATGATGATGAGAAC 149525
Q 914 ----- 914
D 149526 ATTCATGCTGCTTAGAAGTCTCAATGCTATGCTGCTGCTCCCGTACGCTGATC 149585
Q 915 ---CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
D 149586 CGGTGCGCA-GAAGTTCGGGAATCTTGTCTCGCAAGCTGCAAGAGATTAAGCAGAGT 149644
Q 934 Arg-----LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla----- 947
D 149645 TTGCCCAGAACTGTTTGGCGCTGACTTCATGGCGTTGTATGCTGCTGGCTGCTGAG 149704
Q 948 -----LysAspProValLysGluArgArgAlaHisAlaArg-GlnCysLeuValLys 964
D 149705 ACTGAGAGAAATGAGTATGATCAAGAAACCATCCGATCCGTCGCAAGAGTATGATATA 149764
Q 964 s-----AsnIleAsnValArg 969
D 149765 TATTTCAACAGATTGCAAGACSTTGTGCTCACTATGCAAGAAAGGATTAACAAACG 149824
Q 969 GArgGluTyrLeuLys----- 974
D 149825 GCGGGAATATCTCAAGACTGCGCTATGACATGTAAGTGTTTACTTTTGCAATCATATT 149884
Q 975 -----GlnHis-AlaAlaValSerLysLys 983
D 149885 GTGTATTTTAACCATTAATCAATTAATTCACAGCTCCGCAAGCTCAACGAGTCAACAAT 149944
Q 983 LeuLeuSerLeuLeuProGluTyrValAlaProTyrThrIleHisLeuAlaHisAsp 1003
D 149945 CATTAACATACACTGACTACATGCTGCTGCTTCCGCTATTCCTGCTGCTGCACAGATC 150004
Q 1003 roAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTyrP 1023
D 150005 CACGCTTCACCAATACAGAGAGACTACATGACAGTCCGCAAGATGAGAAAGTCCGCTGCT 150064
Q 1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysM 1043
D 150065 TCATTTCTTGAGCCCTCATGATGCCAAGAGAAAGCTTTGTCCATAGCTTATCAACAGCAGC 150124
Q 1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn- 1062
D 150125 TGTGACAGCTGATTAAGCATCGCAGATTCACTTGGGG---TCGGAACAAGCGCAACAAT 150181
Q 1062 ----- 1062
D 150182 ATGTATGCTCACTCAATCATTTTCCAGAGCTTCCCTTAAATTAATTTATATATATAT 150241
Q 1063 -----GluLysLeuTyrThrValCysAspValAlaMetAsnIleLeuMetSerL 1079
D 150242 TTTTTCGTTATCAAGAAATGCGGCGCTGCGATCTTGGCTATGATATTAATCAATTCACA 150301
Q 1079 ys-----SerThrThrTyrSerLeuGluSerProLysAspProv 1092
D 150302 AGTTGACTCATTTGATGGAACAACGAGACACTTTTCCATGATCCGCTG-----G 150349

```
QY 1092 alleProalaArgPheThrGlnProAsp---LysAsnPheSerAsnThrLysAsnT 1111
      |||||
      ::::: |||||
Db 150350 CTTGCCAGAAATGATATTAAGAGCGCTCGGTGGAAATTCACAAACATGACGCT 150409
      |||||
QY 1111 yrlleuProProGluMetLysSerPheThrProGlyLysProLysThrThrAsnVal 1131
      ||::|||
      :::::
Db 150410 ATATACCGCTGAGCGTG-----TATACGCTGGGAGCCAAATCCAGCAGCAAGCTG 150460
      |||||
QY 1131 euGLyAlaValAsnLysProLeu---SerSerAlaGlyLysGlnSerGlnThrLysSer 1150
      |||||
      ::::: |||||
Db 150461 CCGCAACAGC-AATGACACAGCTCGCGCAGCAGTGCTC-CAAAGAGACCGCGCGAC 150518
      |||||
QY 1150 erArg-----MetGluThrValSerAsnAlaSerSer 1161
      ||::|
      ::::: |||||
Db 150519 GTCATATCATGATGATGAAATCCGACGAGTGAAGTCTCTTGAAATATCTGAGAAATG 150578
      |||||
QY 1161 erSerAsnProSerSerProGlyArgLleLysGlyArgLeuAspSerSerGluMetAsp 1181
      |||||
      ::::: |||||
Db 150579 AATCTAAGTATGTGTGAAATTTGANTTTCAGGAGAACAAATCTGTCACACATAC- 150634
      |||||
QY 1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
      |||||
      ::::: |||||
Db 150635 -----GAGCGGCTGACACTACGAGAGCCCATGCCCCAACGAGACGCGCAGGAG 150682
```

Search completed: September 25, 2003, 00:10:10
Job time : 12397 secs